as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations may contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared is such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. For example, a mouthwash may be prepared incorporating the active ingredient in the required amount in an appropriate solvent, such as a sodium borate solution (Dobell's Solution). Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

2. Injectable delivery

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally as described in, *e.g.*, U.S. Patent Nos. 5,543,158; 5,641,515; and 5,399,363. Solutions of the active compounds as free base or pharmacologically

acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Patent No. 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion (*see, e.g.*, Remington *Pharmaceutical Sciences* 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated.

The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human. The preparation of an aqueous composition that contains a protein as an active ingredient is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified.

3. Nasal delivery

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described *e.g.*, in U.S. Patent Nos. 5,756,353 and 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, *J Controlled Release* 52:81-87 (1998)) and lysophosphatidyl-glycerol compounds (*see, e.g.*, U.S. Patent No. 5,725,871) are also well-known in the pharmaceutical arts. Likewise, transmucosal drug delivery in the form of a polytetrafluoroetheylene support matrix is described in U.S. Patent No. 5,780,045.

4. Liposome-, nanocapsule-, and microparticle-mediated delivery

In certain embodiments, the inventors contemplate the use of liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, for the introduction of the compositions of the present invention into suitable host cells. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like.

Such formulations may be preferred for the introduction of pharmaceutically-acceptable formulations of the polypeptides, fusion proteins and nucleic acids disclosed herein. The formation and use of liposomes is generally known to those of skill in the art (*see, e.g.*, Couvreur *et al.*, *FEBS Lett.* *84*(2):323-326 (1977); Couvreur (1988); Lasic, *Trends Biotechnol.* *16*(7):307-321 (1998); which describes the use of liposomes and nanocapsules in the targeted antibiotic therapy for intracellular bacterial

infections and diseases). Recently, liposomes were developed with improved serum stability and circulation half-times (Gabizon and Papahadjopoulos, *Proc Natl Acad Sci U S A. 85*(18):6949-6953 (1988); Allen and Choun (1987); U.S. Patent No. 5,741,516). Further, various methods of liposome and liposome like preparations as potential drug carriers have been reviewed (Takakura, *Nippon Rinsho 56*(3):691-695 (1998); Chandran *et al.*, *Indian J Exp Biol. 35*(8):801-809 (1997); Margalit, *Crit Rev Ther Drug Carrier Syst. 12*(2-3):233-261 (1995); U.S. Patent Nos. 5,567,434; 5,552,157; 5,565,213; 5,738,868; and 5,795,587).

Liposomes have been used successfully with a number of cell types that are normally resistant to transfection by other procedures including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, *J Biol Chem. 265*(27):16337-16342 (1990); Muller *et al.*, *DNA Cell Biol. 9*(3):221-229 (1990)). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, drugs (Heath and Martin, *Chem Phys Lipids 40*(2-4):347-358 (1986); Heath *et al.*, *Biochim Biophys Acta. 862*(1):72-80 (1986); Balazsovits *et al.*, *Cancer Chemother Pharmacol. 23*(2):81-6. (1989); Fresta and Puglisi, *J. Drug Target 4*(2):95-101 (1996)), radiotherapeutic agents (Pikul *et al.*, *Arch Surg. 122*(12):1417-1420 (1987)), enzymes (Imaizumi *et al.*, *Stroke 21*(9):1312-1317 (1990); Imaizumi *et al.*, *Acta Neurochir Suppl (Wien) 51*:236-238 (1990)), viruses (Faller and Baltimore, *J Virol. 49*(1):269-272 (1984)), transcription factors and allosteric effectors (Nicolau and Gersonde, *Naturwissenschaften 66*(11):563-566 (1979)) into a variety of cultured cell lines and animals. In addition, several successful clinical trails examining the effectiveness of liposome-mediated drug delivery have been completed (Lopez-Berestein *et al.*, *J Infect Dis. 151*(4):704-710 (1985); Lopez-Berestein *et al.*, *Cancer Drug Deliv. 2*(3):183-189 (1985); Coune, *Infection 16*(3):141-147 (1988); Sculier *et al.*, *Eur. J. Cancer Clin. Oncol. 24*(3):527-38 (1988)). Furthermore, several studies suggest that the use of liposomes is not associated with autoimmune responses, toxicity or gonadal localization after systemic delivery (Mori and Fukatsu, *Epilepsia 33*(6):994-1000 (1992)).

Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μm.

Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 Å, containing an aqueous solution in the core.

Liposomes bear resemblance to cellular membranes and are contemplated for use in connection with the present invention as carriers for the peptide compositions. They are widely suitable as both water- and lipid-soluble substances can be entrapped, *i.e.* in the aqueous spaces and within the bilayer itself, respectively. It is possible that the drug-bearing liposomes may even be employed for site-specific delivery of active agents by selectively modifying the liposomal formulation.

In addition to the teachings of Couvreur *et al.* (1977), *supra*; Couvreur *et al.* (1988), *supra*), the following information may be utilized in generating liposomal formulations. Phospholipids can form a variety of structures other than liposomes when dispersed in water, depending on the molar ratio of lipid to water. At low ratios the liposome is the preferred structure. The physical characteristics of liposomes depend on pH, ionic strength and the presence of divalent cations. Liposomes can show low permeability to ionic and polar substances, but at elevated temperatures undergo a phase transition which markedly alters their permeability. The phase transition involves a change from a closely packed, ordered structure, known as the gel state, to a loosely packed, less-ordered structure, known as the fluid state. This occurs at a characteristic phase-transition temperature and results in an increase in permeability to ions, sugars and drugs.

In addition to temperature, exposure to proteins can alter the permeability of liposomes. Certain soluble proteins, such as cytochrome c, bind, deform and penetrate the bilayer, thereby causing changes in permeability. Cholesterol inhibits this penetration of proteins, apparently by packing the phospholipids more tightly. It is contemplated that the most useful liposome formations for antibiotic and inhibitor delivery will contain cholesterol.

The ability to trap solutes varies between different types of liposomes. For example, MLVs are moderately efficient at trapping solutes, but SUVs are extremely inefficient. SUVs offer the advantage of homogeneity and reproducibility in size distribution, however, and a compromise between size and trapping efficiency is offered by large unilamellar vesicles (LUVs). These are prepared by ether evaporation and are three to four times more efficient at solute entrapment than MLVs.

In addition to liposome characteristics, an important determinant in entrapping compounds is the physicochemical properties of the compound itself. Polar compounds are trapped in the aqueous spaces and nonpolar compounds bind to the lipid bilayer of the vesicle. Polar compounds are released through permeation or when the bilayer is broken, but nonpolar compounds remain affiliated with the bilayer unless it is disrupted by temperature or exposure to lipoproteins. Both types show maximum efflux rates at the phase transition temperature.

Liposomes interact with cells *via* four different mechanisms: endocytosis by phagocytic cells of the reticuloendothelial system such as macrophages and neutrophils; adsorption to the cell surface, either by nonspecific weak hydrophobic or electrostatic forces, or by specific interactions with cell-surface components; fusion with the plasma cell membrane by insertion of the lipid bilayer of the liposome into the plasma membrane, with simultaneous release of liposomal contents into the cytoplasm; and by transfer of liposomal lipids to cellular or subcellular membranes, or vice versa, without any association of the liposome contents. It often is difficult to determine which mechanism is operative and more than one may operate at the same time.

The fate and disposition of intravenously injected liposomes depend on their physical properties, such as size, fluidity, and surface charge. They may persist in tissues for hours or days, depending on their composition, and half lives in the blood range from minutes to several hours. Larger liposomes, such as MLVs and LUVs, are taken up rapidly by phagocytic cells of the reticuloendothelial system, but physiology of the circulatory system restrains the exit of such large species at most sites. They can exit only in places where large openings or pores exist in the capillary endothelium, such as the sinusoids of the liver or spleen. Thus, these organs are the predominate site of uptake. On the other hand, SUVs show a broader tissue distribution but still are sequestered highly in the liver and spleen. In general, this *in vivo* behavior limits the potential targeting of liposomes to only those organs and tissues accessible to their large size. These include the blood, liver, spleen, bone marrow, and lymphoid organs.

Targeting is generally not a limitation in terms of the present invention. However, should specific targeting be desired, methods are available for this to be accomplished. Antibodies may be used to bind to the liposome surface and to direct the antibody and its drug contents to specific antigenic receptors located on a particular cell-

type surface. Carbohydrate determinants (glycoprotein or glycolipid cell-surface components that play a role in cell-cell recognition, interaction and adhesion) may also be used as recognition sites as they have potential in directing liposomes to particular cell types. Mostly, it is contemplated that intravenous injection of liposomal preparations would be used, but other routes of administration are also conceivable.

Alternatively, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (Henry-Michelland *et al.* (1987); Quintanar-Guerrero *et al.*, *Pharm Res. 15*(7):1056-1062 (1998); Douglas *et al.*, *Crit. Rev. Ther. Drug Carrier Syst. 3*(3):233-261 (1987)). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present invention. Such particles may be are easily made, as described (Couvreur *et al.*, *J. Pharm. Sci. 69*(2):199-202 (1980); Couvreur *et al.*, (1988), *supra*; zur Muhlen *et al.*, *Eur. J. Pharm. Biopharm. 45*(2):149-155 (1998); Zambaux *et al.*, *J. Controlled Release 50*(1-3):31-40 (1998); Pinto-Alphandry *et al.* (1995); and U.S. Patent No. 5,145,684).

B. Vaccines

In certain preferred embodiments of the present invention, vaccines are provided. The vaccines will generally comprise one or more pharmaceutical compositions, such as those discussed above, in combination with a non-specific immune response enhancer. A non-specific immune response enhancer may be any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. Examples of non-specific immune response enhancers include adjuvants, biodegradable microspheres (*e.g.*, polylactic galactide) and liposomes (into which the compound is incorporated; *see e.g.*, Fullerton, U.S. Patent No. 4,235,877). Vaccine preparation is generally described in, for example, Powell and Newman, eds., "*Vaccine Design (the subunit and adjuvant approach)*," Plenum Press (NY, 1995). Vaccines may be designed to generate antibody immunity and/or cellular immunity such as that arising from CTL or CD4+ T cells.

Pharmaceutical compositions and vaccines within the scope of the present invention may also contain other compounds, which may be biologically active or inactive. For example, one or more immunogenic portions of other *Mycobacterium* antigens may be present, either incorporated into a fusion polypeptide or as a separate compound, within the composition or vaccine. Polypeptides may, but need not, be conjugated to other macromolecules as described, for example, within U.S. Patent Nos. 4,372,945 and 4,474,757. Pharmaceutical compositions and vaccines may generally be used for prophylactic and therapeutic purposes.

Illustrative vaccines may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. Such a polynucleotide may comprise DNA, RNA, a modified nucleic acid or a DNA/RNA hybrid. As noted above, the nucleic acid may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacteria and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems 15*:143-198 (1998), and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (*e.g.*, vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch *et al.*, *Proc. Natl. Acad. Sci. USA 86*:317-321 (1989); Flexner *et al.*, *Ann. N.Y. Acad. Sci. 569*:86-103 (1989); Flexner *et al.*, *Vaccine 8*:17-21 (1990); U.S. Patent Nos. 4,603,112; 4,769,330; and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques 6*:616-627 (1988); Rosenfeld *et al.*, *Science 252*:431-434 (1991); Kolls *et al.*, *Proc. Natl. Acad. Sci. USA 91*:215-219 (1994); Kass-Eisler *et al.*, *Proc. Natl. Acad. Sci. USA 90*:11498-11502 (1993); Guzman *et al.*, *Circulation 88*:2838-2848 (1993); and Guzman *et al.*, *Cir. Res. 73*:1202-1207 (1993). Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art.

The DNA may also be "naked," as described, for example, in Ulmer *et al.*, *Science 259*:1745-1749 (1993) and reviewed by Cohen, *Science 259*:1691-1692 (1993). The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells. It will be apparent that a vaccine may comprise both a polynucleotide and a polypeptide component. Such vaccines may provide for an enhanced immune response.

In a related aspect, a DNA vaccine as described *supra* may be administered simultaneously with or sequentially to either a polypeptide of the present invention or a known *Mycobacterium* antigen, such as the 38 kD antigen described above For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described *supra*, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

It will be apparent that a vaccine may contain pharmaceutically acceptable salts of the polynucleotides and polypeptides provided herein. Such salts may be prepared from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (*e.g.*, sodium, potassium, lithium, ammonium, calcium and magnesium salts).

While any suitable carrier known to those of ordinary skill in the art may be employed in the vaccine compositions of this invention, the type of carrier will vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, intravenous, intracranial, intraperitoneal, subcutaneous or intramuscular administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (*e.g.*, polylactate polyglycolate) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344 and 5,942,252. One may also employ a carrier comprising the particulate-protein

complexes described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

Such compositions may also comprise buffers (*e.g.*, neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate. Compounds may also be encapsulated within liposomes using well known technology.

Any of a variety of immunostimulants may be employed in the vaccines of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bortadella pertussis* or *Mycobacterium* species or *Mycobacterium* derived proteins. For example, delipidated, deglycolipidated *M. vaccae* ("pVac") can be used. In another embodiment, BCG is used as an adjuvant. In addition, the vaccine can be administered to a subject previously exposed to BCG. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 and derivatives thereof (SmithKline Beecham, Philadelphia, PA); CWS, TDM, Leif, aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2, -7, or -12, may also be used as adjuvants.

Within the vaccines provided herein, the adjuvant composition is preferably designed to induce an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN-γ, TNFα, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided

herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann & Coffman, *Ann. Rev. Immunol.* 7:145-173 (1989).

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, WA; *see* US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato *et al.*, *Science* 273:352 (1996). Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins . Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamelar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol® to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL® adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 as disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2, AS2', AS2,'' SBAS-4, or SBAS6, available from SmithKline Beecham, Rixensart, Belgium), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula (I): $HO(CH_2CH_2O)_n$-A-R,

wherein, n is 1-50, A is a bond or –C(O)-, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein *n* is between 1 and 50, preferably 4-24, most preferably 9; the *R* component is C_{1-50}, preferably C_4-C_{20} alkyl and most preferably C_{12} alkyl, and *A* is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-steoryl ether, polyoxyethylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl

ether are described in the Merck index (12^{th} edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

Any vaccine provided herein may be prepared using well known methods that result in a combination of antigen, immune response enhancer and a suitable carrier or excipient. The compositions described herein may be administered as part of a sustained release formulation (*i.e.,* a formulation such as a capsule, sponge or gel (composed of polysaccharides, for example) that effects a slow release of compound following administration). Such formulations may generally be prepared using well known technology (*see, e.g.*, Coombes *et al., Vaccine 14*:1429-1438 (1996)) and administered by, for example, oral, rectal or subcutaneous implantation, or by implantation at the desired target site. Sustained-release formulations may contain a polypeptide, polynucleotide or antibody dispersed in a carrier matrix and/or contained within a reservoir surrounded by a rate controlling membrane.

Carriers for use within such formulations are biocompatible, and may also be biodegradable; preferably the formulation provides a relatively constant level of active component release. Such carriers include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

C. Delivery vehicles

Any of a variety of delivery vehicles may be employed within pharmaceutical compositions and vaccines to facilitate production of an antigen-specific immune response that targets infected cells. Delivery vehicles include antigen presenting

cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified, *e.g.*, to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response and/or to be immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature 392*:245-251 (1998)) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic immunity (*see* Timmerman and Levy, *Ann. Rev. Med. 50*:507-529 (1999)). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel *et al.*, *Nature Med. 4*:594-600 (1998)).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα, CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible

intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.*, CD54 and CD11) and costimulatory molecules (*e.g.*, CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide encoding a *Mycobacterium* antigen (or portion or other variant thereof) such that the *Mycobacterium* polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a composition or vaccine comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in, *e.g.*, WO 97/24447, or the gene gun approach described by Mahvi *et al.*, *Immunology and cell Biology* 75:456-460 (1997). Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the *Mycobacterium* polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

D. Therapeutic applications of the compositions of the invention

In further aspects of the present invention, the compositions described *supra* may be used for immunotherapy of *Mycobacterium* infection, and in particular tuberculosis. Within such methods, pharmaceutical compositions and vaccines are typically administered to a patient to either prevent the development of *Mycobacterium* infection or to treat a patient afflicted with *Mycobacterium* infection. *Mycobacterium* infection may be diagnosed using criteria generally accepted in the art, such as, *e.g.*, in

the case of tuberculosis, fever, acute inflammation of the lung and/or non-productive cough. Pharmaceutical compositions and vaccines may be administered either prior to or following a treatment such as administration of conventional drugs. Administration may be by any suitable route, including, *e.g.*, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, oral, *etc.*

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against *Mycobacterium* infection with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established *Mycobacterium*-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate anti-*Mycobacterium* infection effects and do not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as $CD8^+$ cytotoxic T lymphocytes and $CD4^+$ T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide of the invention. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In

particular, antigen-presenting cells, such as dendritic, macrophage or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (*see*, for example, Cheever *et al.*, *Immunological Reviews 157*:177, (1997)).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by, *e.g.*, injection, intranasal or oral administration.

E. Formulation and administration

Vaccines and pharmaceutical compositions may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are preferably hermetically sealed to preserve sterility of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a vaccine or pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

Routes and frequency of administration, as well as dosage, may vary from individual to individual and may parallel those currently being employed in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered, *e.g.*, by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described *supra*, is capable of raising an

immune response in an immunized patient sufficient to protect the patient from *Mycobacterium* infection for at least 1-2 years. When used for a therapeutic purpose, a suitable dose is the amount that is capable of raising and immune response in a patient that is sufficient to obtain an improved clinical outcome (*e.g.*, more frequent cure) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a *Mycobacterium* protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

In general, the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 µg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 ml to about 5 ml.

F. Diagnostic kits

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a *Mycobacterium* antigen. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a *Mycobacterium* antigen in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a *Mycobacterium* antigen. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent

or container to facilitate the detection of a polynucleotide encoding a *Mycobacterium* antigen.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

VIII. EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* POLYPEPTIDES USING CD4+ T CELL LINES GENERATED FROM HUMAN PBMC

M. tuberculosis antigens of the present invention were isolated by expression cloning of cDNA libraries of *M. tuberculosis* strains H37Rv and Erdman essentially as described by Sanderson *et al.* (*J. Exp. Med.*, *182*:1751-1757 (1995)) and were shown to induce PBMC proliferation and IFN-γ in an immunoreactive T cell line.

Two CD4+ T cell lines, referred to as DC-4 and DC-5, were generated against dendritic cells infected with *M. tuberculosis*. Specifically, dendritic cells were prepared from adherent PBMC from a single donor and subsequently infected with tuberculosis. Lymphocytes from the same donor were cultured under limiting dilution conditions with the infected dendritic cells to generate the CD4+ T cell lines DC-4 and DC-5. These cell lines were shown to react with crude soluble proteins from *M. tuberculosis* but not with Tb38-1. Limiting dilution conditions were employed to obtain a third CD4+ T cell line, referred to as DC-6, which was shown to react with both crude soluble proteins and Tb38-1.

Genomic DNA was isolated from the *M. tuberculosis* strains H37Rv and Erdman and used to construct expression libraries in the vector pBSK(-) using the

Lambda ZAP expression system (Stratagene, La Jolla, CA). These libraries were transformed into *E. coli*, pools of induced *E. coli* cultures were incubated with dendritic cells, and the ability of the resulting incubated dendritic cells to stimulate cell proliferation and IFN-γ production in the CD4+ T cell line DC-6 was examined as described below in Example 2. Positive pools were fractionated and re-tested until pure *M. tuberculosis* clones were obtained.

Nineteen clones were isolated, of which nine were found to contain the previously identified *M. tuberculosis* antigens TbH-9 and Tb38-1, disclosed in U.S. Patent Application No. 08/533,634. The determined cDNA sequences for the remaining ten clones (hereinafter referred to as Tb224, Tb636, Tb424, Tb436, Tb398, Tb508, Tb44l, Tb475, Tb488 and Tb465) are provided in SEQ ID NO:1-10, respectively. The corresponding predicted amino acid sequences for Tb224 and Tb636 are provided in SEQ ID NO:13 and 14, respectively. The open reading frames for these two antigens were found to show some homology to TbH-9. Tb224 and Tb636 were also found to be overlapping clones.

Tb424, Tb436, Tb398, Tb508, Tb441, Tb475, Tb488 and Tb465 were each found to contain two small open reading frames (referred to as ORF-1 and ORF-2) or truncated forms thereof, with minor variations in ORF-1 and ORF-2 being found for each clone. The predicted amino acid sequences of ORF-1 and ORF-2 for Tb424, Tb436, Tb398, Tb508, Tb441, Tb475, Tb488 and Tb465 are provided in SEQ ID NO:16 and 17, 18 and 19, 20 and 21, 22 and 23, 24 and 25, 26 and 27, 28 and 29, and 30 and 31, respectively. In addition, clones Tb424 and Tb436 were found to contain a third apparent open reading frame, referred to as ORF-U. The predicted amino acid sequences of ORF-U for Tb424 and Tb436 are provided in SEQ ID NO:32 and 33, respectively. Tb424 and Tb436 were found to be either overlapping clones or recently duplicated/transposed copies. Similarly Tb398, Tb508 and Tb465 were found to be either overlapping clones or recently duplicated/transposed copies, as were Tb475 and Tb488.

These sequences were compared with known sequences in publicly available sequence databases using the BLASTN system. No homologies to the antigens Tb224 and Tb431 were found. Tb636 was found to be 100% identical to a cosmid previously identified in *M. tuberculosis*. Similarly, Tb508, Tb488, Tb398, Tb424, Tb436, Tb441, Tb465 and Tb475 were found to show homology to known *M. tuberculosis*

cosmids. In addition, Tb488 was found to have 100% homology to *M. tuberculosis* topoisomerase I.

Seventeen overlapping peptides to the open reading frames ORF-1 (referred to as 1-1 - 1-17; SEQ ID NO:34-50, respectively) and thirty overlapping peptides to the open reading frame ORF-2 (referred to as 2-1 - 2-30, SEQ ID NO:51-80, respectively) were synthesized using the procedure described below in Example 4.

The ability of the synthetic peptides and of recombinant ORF-1 and ORF-2 to induce T cell proliferation and IFN-γ production in PBMC from PPD-positive donors was assayed as described below in Example 2. Figs. 1A-B and 2A-B illustrate stimulation of T cell proliferation and IFN-γ by recombinant ORF-2 and the synthetic peptides 2-1 - 2-16 for two donors, referred to as D7 and D160, respectively. Recombinant ORF-2 (referred to as MTI) stimulated T cell proliferation and IFN-γ production in PBMC from both donors. The amount of PBMC stimulation seen with the individual synthetic peptides varied with each donor, indicating that each donor recognizes different epitopes on ORF-2. The proteins encoded by ORF-1, ORF-2 and ORF-U were subsequently named MTS, MTI and MSF, respectively.

Eighteen overlapping peptides to the sequence of MSF (referred to as MSF-1- MSF-18; SEQ ID NO:84-101, respectively) were synthesized and their ability to stimulate T cell proliferation and IFN-γ production in a CD4+ T cell line generated against *M. tuberculosis* culture filtrate was examined as described below. The peptides referred to as MSF-12 and MSF-13 (SEQ ID NO:95 and 96, respectively) were found to show the highest levels of reactivity.

Two overlapping peptides (SEQ ID NO:81 and 82) to the open reading frame of Tb224 were synthesized and shown to induce T cell proliferation and IFN-γ production in PBMC from PPD-positive donors.

Two CD4+ T cell lines from different donors were generated against *M. tuberculosis* infected dendritic cells using the above methodology. Screening of the *M. tuberculosis* cDNA expression library described above using this cell line, resulted in the isolation of two clones referred to as Tb867 and Tb391. The determined cDNA sequence for Tb867 (SEQ ID NO:102) was found to be identical to the previously isolated *M. tuberculosis* cosmid SCY22G10, with the candidate reactive open reading frame encoding

a 750 amino acid *M. tuberculosis* protein kinase. Comparison of the determined cDNA sequence for Tb391 (SEQ ID NO:103) with those in publicly available sequence databases revealed no significant homologies to known sequences.

In further studies, CD4+ T cell lines were generated against *M. tuberculosis* culture filtrate, essentially as outlined above, and used to screen the *M. tuberculosis* Erdman cDNA expression library described above. Five reactive clones, referred to as Tb431, Tb472, Tb470, Tb838 and Tb962 were isolated. The determined cDNA sequences for Tb431, Tb472, Tb470, and Tb838 are provided in SEQ ID NO:11, 12, 104 and 105, respectively, with the determined cDNA sequences for Tb962 being provided in SEQ ID NO:106 and 107. The corresponding predicted amino acid sequence for Tb431 is provided in SEQ ID NO:15.

Subsequent studies led to the isolation of a full-length cDNA sequence for Tb472 (SEQ ID NO:108). Overlapping peptides were synthesized and used to identify the reactive open reading frame. The predicted amino acid sequence for the protein encoded by Tb472 (referred to as MSL) is provided in SEQ ID NO:109. Comparison of the sequences for Tb472 and MSL with those in publicly available sequence databases as described above, revealed no homologies to known sequences. Fifteen overlapping peptides to the sequence of MSL (referred to as MSL-1 - MSL-15; SEQ ID NO:110-124, respectively) were synthesized and their ability to stimulate T cell proliferation and IFN-γ production in a CD4+ T cell line generated against *M. tuberculosis* culture filtrate was examined as described below. The peptides referred to as MSL-10 (SEQ ID NO:119) and MSL-11 (SEQ ID NO:120) were found to show the highest level of reactivity. Comparison of the determined cDNA sequence for Tb838 with those in publicly available sequence databases revealed identity to the previously isolated *M. tuberculosis* cosmid SCY07H7. Comparison of the determined cDNA sequences for the clone Tb962 with those in publicly available sequence databases revealed some homology to two previously identified *M. tuberculosis* cosmids, one encoding a portion of bactoferritin. However, recombinant bactoferritin was not found to be reactive with the T cell line used to isolate Tb962.

The clone Tb470, described above, was used to recover a full-length open reading frame (SEQ ID NO:125) that showed homology with TbH9 and was found to encode a 40 kDa antigen, referred to as Mtb40. The determined amino acid sequence for

Mtb40 is provided in SEQ ID NO:126. Similarly, subsequent studies led to the isolation of the full-length cDNA sequence for Tb431, provided in SEQ ID NO:83, which was also determined to contain an open reading frame encoding Mtb40. Tb470 and Tb431 were also found to contain a potential open reading frame encoding a U-ORF-like antigen.

Screening of an *M. tuberculosis* Erdman cDNA expression library with multiple CD4+ T cell lines generated against *M. tuberculosis* culture filtrate, resulted in the isolation of three clones, referred to as Tb366, Tb433 and Tb439. The determined cDNA sequences for Tb366, Tb433 and Tb439 are provided in SEQ ID NO:127, 128 and 129, respectively. Comparison of these sequences with those in publicly available sequence databases revealed no significant homologies to Tb366. Tb433 was found to show some homology to the previously identified *M. tuberculosis* antigen MPT83. Tb439 was found to show 100% identity to the previously isolated *M. tuberculosis* cosmid SCY02B10.

A CD4+ T cell line was generated against *M. tuberculosis* PPD, essentially described above, and used to screen the above *M. tuberculosis* Erdman cDNA expression library. One reactive clone (referred to as Tb372) was isolated, with the determined cDNA sequences being provided in SEQ ID NO:130 and 131. Comparison of these sequences with those in publicly available sequence databases revealed no significant homologies.

In further studies, screening of an *M. tuberculosis* cDNA expression library with a CD4+ T cell line generated against dendritic cells that had been infected with tuberculosis for 8 days, as described above, led to the isolation of two clones referred to as Tb390R5C6 and Tb390R2C11. The determined cDNA sequence for Tb390R5C6 is provided in SEQ ID NO:132, with the determined cDNA sequences for Tb390R2C11 being provided in SEQ ID NO:133 and 134. Tb390R5C6 was found to show 100% identity to a previously identified *M. tuberculosis* cosmid.

In subsequent studies, the methodology described above was used to screen an *M. tuberculosis* genomic DNA library prepared as follows. Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The phage library (referred to

as the Erd λScreen library) was amplified and a portion was converted into a plasmid expression library by an autosubcloning mechanism using the *E. coli* strain BM25.8 (Novagen). Plasmid DNA was purified from BM25.8 cultures containing the pSCREEN recombinants and used to transform competent cells of the expressing host strain BL21(DE3)pLysS. Transformed cells were aliquoted into 96 well microtiter plates with each well containing a pool size of approximately 50 colonies. Replica plates of the 96 well plasmid library format were induced with IPTG to allow recombinant protein expression. Following induction, the plates were centrifuged to pellet the *E. coli* which was used directly in T cell expression cloning of a CD4+ T cell line prepared from a PPD-positive donor (donor 160) as described above. Pools containing *E. coli* expressing *M. tuberculosis* T cell antigens were subsequently broken down into individual colonies and reassayed in a similar fashion to identify positive hits.

Screening of the T cell line from donor 160 with one 96 well plate of the Erd λScreen library provided a total of nine positive hits. Previous experiments on the screening of the pBSK library described above with T cells from donor 160 suggested that most or all of the positive clones would be TbH-9, Tb38-1 or MTI (disclosed in U.S. Patent Application No. 08/533,634) or variants thereof. However, Southern analysis revealed that only three wells hybridized with a mixed probe of TbH-9, Tb38-1 and MTI. Of the remaining six positive wells, two were found to be identical. The determined 5' cDNA sequences for two of the isolated clones (referred to as Y1-26C1 and Y1-86C11) are provided in SEQ ID NO:135 and 136, respectively. The full length cDNA sequence for the isolated clone referred to as hTcc#1 is provided in SEQ ID NO:137, with the corresponding predicted amino acid sequence being provided in SEQ ID NO:138. Comparison of the sequences of hTcc#1 to those in publicly available sequence databases as described above, revealed some homology to the previously isolated *M. tuberculosis* cosmid MTCY07H7B.06.

EXAMPLE 2

INDUCTION OF T CELL PROLIFERATION AND INTERFERON-γ PRODUCTION BY *M. TUBERCULOSIS* ANTIGENS

The ability of recombinant *M. tuberculosis* antigens to induce T-cell proliferation and interferon-γ production may be determined as follows.

Proteins may be induced by IPTG and purified by gel elution, as described in Skeiky *et al.*, *J. Exp. Med.* *181*:1527-1537 (1995). The purified polypeptides are then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells are known to proliferate in response to PPD are cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 µg/ml gentamicin. Purified polypeptides are added in duplicate at concentrations of 0.5 to 10 µg/ml. After six days of culture in 96-well round-bottom plates in a volume of 200 µl, 50 µl of medium is removed from each well for determination of IFN-γ levels, as described below. The plates are then pulsed with 1 µCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that result in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone are considered positive.

IFN-γ is measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates are coated with a mouse monoclonal antibody directed to human IFN-γ (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells are then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates are washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates are incubated overnight at room temperature. The plates are again washed and a polyclonal rabbit anti-human IFN-γ serum diluted 1:3000 in PBS/10% normal goat serum is added to each well. The plates are then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical So., St. Louis, MO) is added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates are washed and TMB substrate added. The reaction is stopped after 20 min with 1 N sulfuric acid. Optical density is determined at 450 nm using 570 nm as a reference wavelength. Fractions that result in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, are considered positive.

EXAMPLE 3

PURIFICATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* POLYPEPTIDES USING CD4+ T CELL LINES GENERATED FROM A MOUSE *M. TUBERCULOSIS* MODEL

Infection of C57BL/6 mice with *M. tuberculosis* results in the development of a progressive disease for approximately 2-3 weeks. The disease progression is then halted as a consequence of the emergence of a strong protective T cell-mediated immune response. This infection model was used to generate T cell lines capable of recognizing protective *M. tuberculosis* antigens.

Specifically, spleen cells were obtained from C57BL/6 mice infected with *M. tuberculosis* for 28 days and used to raise specific anti-*M. tuberculosis* T cell lines as described above. The resulting CD4+ T cell lines, in conjunction with normal antigen presenting (spleen) cells from C57BL/6 mice were used to screen the *M. tuberculosis* Erd λScreen library described above. One of the reactive library pools, which was found to be highly stimulatory of the T cells, was selected and the corresponding active clone (referred to as Y288C10) was isolated.

Sequencing of the clone Y288C10 revealed that it contains two potential genes, in tandem. The determined cDNA sequences for these two genes (referred to as mTCC#1 and mTCC#2) are provided in SEQ ID NO:139 and 140, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO:141 and 142, respectively. Comparison of these sequences with those in publicly available sequence databases revealed identity to unknown sequences previously found within the *M. tuberculosis* cosmid MTY21C12. The predicted amino acid sequences of mTCC#1 and mTCC#2 were found to show some homology to previously identified members of the TbH9 protein family, discussed above.

EXAMPLE 4

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following

cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

EXAMPLE 5

USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

The diagnostic properties of representative *M. tuberculosis* antigens may be determined by examining the reactivity of antigens with sera from tuberculosis-infected patients and from normal donors as described below.

Assays are performed in 96-well plates coated with 200 ng antigen diluted to 50 µl in carbonate coating buffer, pH 9.6. The wells are coated overnight at 4°C (or 2 hours at 37°C). The plate contents are then removed and the wells are blocked for 2 hours with 200 µl of PBS/1% BSA. After the blocking step, the wells are washed five times with PBS/0.1% Tween 20™. 50 µl sera, diluted 1:100 in PBS/0.1% Tween 20/0.1% BSA, is then added to each well and incubated for 30 minutes at room temperature. The plates are washed again five times with PBS/0.1% Tween 20™.

The enzyme conjugate (horseradish peroxidase - Protein A, Zymed, San Francisco, CA) is then 1:10,000 in PBS/0.1% Tween20™/0.1% BSA, and 50 µl of the diluted conjugate is added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells are washed five times with PBS/0.1% Tween 20™. 100 µl of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) is added, undiluted, and incubated for about 15 minutes. The reaction is stopped with the addition of 100 µl of 1 NH_2SO_4 to each well, and the plates are read at 450 nm.

EXAMPLE 6

MURINE T CELL EXPRESSION CLONING OF AN MTB ANTIGEN ASSOCIATED WITH THE CONTROL OF TB INFECTION

Genomic DNA form *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, blunt ended with Klenow polymerase and followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector predigested with EcoRI (Novagen, Madison, WI) and packaged in vitro using the PhageMaker extract (Novagen, Madison, WI). The phage library (Erd Screen) was amplified and a portion converted into a plasmid expression library (pScreen) by autosubcloning using the *E. coli* host strain BM25.8 as suggested by the manufacturer (Novagen, Madison, WI). Plasmid DNA was purified from BM25.8 cultures containing pScreen recombinants and used to transform competent cells of the expressing host strain BL21(DE3)pLysS. Transformed cells were aliquoted into 96 well micro titer plates with each well containing a pool size of ~50 colonies. Replica plates of the 96 well plasmid library format were induced with IPTG to allow recombinant protein expression. Following induction, the plates were centrifuged to pellet the *E. coli* and the bacterial pellet was resuspended in 200 µl of 1X PBS. The general principle is based on the direct recognition by the T cells of the antigens presented by antigen presenting cells that have internalized a library of *E. coli*-containing expressed recombinant antigens. The *M. tuberculosis* library was initially divided in pools containing approximately 50-100 transformants/ml distributed in 96-well microtiter plates and stored in a replica plate manner. Adherent spleen cells were fed with the *E. coli* pools and incubated for processing for 2 h. After washing the adherent cells were exposed to specific T cell lines in the presence of gentamycin (50 µg/ml) to inhibit the bacterial growth. T cell recognition of pool containing *M. tuberculosis* antigens was then detected by proliferation (3H thymidine incorporation). Wells that scored positive were then broken down using the same protocol until a single clone was detected. The gene was then sequenced, sub-cloned, expressed and the recombinant protein evaluated. Nucleotide sequence comparison of the 0.6 kb insert of clone mTTC#3 with the GenBank database revealed that it is comprised of the amino terminal portion of gene MTV014.03c (locus MTV014; accession # e1248750) of the Mtb H37Rv strain. The full length nucleotide sequence of mTTC#3 (SEQ ID NO:145) is a 1.86 kb fragment comprising the entire ORF with a

predicted molecular weight of ~57 kDa (SEQ ID NO:146). Thus, to maintain consistency with our nomenclature, mTTC#3 is referred to hereafter as MTB57. The full length coding portion of mTTC#3 (MTB57) was PCR amplified using the following primer pairs: 5'(5' –CAA TTA CAT ATG CAT CAC CAT CAC CAT CAC ATG AAT TAT TCG GTG TTG CCG (SEQ ID NO:147)) and 3' (5' –CAA TTA AAG CTT TTA GGG CTG ACC GAA GAA GCC (SEQ ID NO:148))h3. The full length nucleic acid coding sequence of mTTC#3 and the corresponding predicted amino acid sequence are provided in Figures 3 and 4, respectively.

EXAMPLE 7

IDENTIFICATION OF MYCOBACTERIUM TUBERCULOSIS ANTIGENS EXCRETED IN URINE OF INFECTED MICE

Antigen were prepared by infecting intravenously C57BL/6 mice with 4.10^{7} colony forming units (CFU) of *M. tuberculosis*. 14 days later the animals were bled and their urine was collected in microfuge tubes. Sera were obtained at room temperature. Both sera and urine were centrifuged at 10,000 g for 15 minutes followed by filtration in 0.2u sterile membranes.

Antibodies were produced against the antigens by immunizing normal C57BL/6 mice with either the sera or the urine from the *M. tuberculosis* infected C57BL/6 mice. The adjuvant used was incomplete Freund's adjuvant (IFA). Immunization was carried out according to the following protocol: on day 1, mice were injected in the footpad or in the base of the tail with a mix containing 100 µl of either serum or urine and 100 µl of IFA; on day 14, a mix containing 100 µl of either serum or urine and 100 µl of IFA was injected intraperitoneally to the mice; finally on day 28, either 200 µl of serum or 50 µl of urine were injected to the mice intraperitoneally. By using syngeneic mice for the antibody production, only antibodies specific for foreign antigens present in the blood circulation or urine of the C57BL/6 mice, *i.e.*, *M. tuberculosis* antigens, are generated. On day 35, 100 µl of blood were collected by eye-bleeding the immunized mice. ELISA assays were performed with the obtained sera using a *M. tuberculosis* crude lysate. The ELISA experiments revealed that all the mice immunized with either sera or urine from infected donors produced anti-*M. tuberculosis* antibodies in titers varying from 1/40 to 1/320. No anti-*M. tuberculosis* antibodies were found in the sera obtained from the mice before the immunizations.

The antiserum made against the proteins excreted in the urine was used to screen a Mtb expression library prepared in the lambda screen phage expression system. Positive clones were purified and their corresponding inserts sequenced. These inserts were named P1, 2, 3, 4, 6, 7, 8, 9, 10, 11 and 12 (SEQ ID NO:149-159).

EXAMPLE 8

IDENTIFICATION OF MYCOBACTERIUM TUBERCULOSIS ANTIGENS USING CD4+ T CELL EXPRESSION CLONING

Expression screening using a number of T cell lines generated from healthy PPD-positive individuals has been employed to identify *M. tuberculosis* clones encoding reactive antigens. Pools of *M. tuberculosis* recombinant clones (expressed in *E. coli*) were fed to dendritic cells. Autologous T cell lines were incubated with the dendritic cells and proliferation and INF-gamma production was measured. Reactive pools were fractionated and re-tested until pure *M. tuberculosis* clones were achieved. This approach allows for direct screening for T cell antigens. A related approach has been used to identify *Listeria monocytogenes* antigens (see *J. Exp. Med.* 182:1751-1757 (1995).

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. An isolated polypeptide comprising an amino acid sequence of SEQ ID NO:146, 161, or 163, or an amino acid sequence comprising an immunogenic portion of an amino acid sequence of SEQ ID NO:146, 161, or 163.

2. An isolated polypeptide, wherein said polypeptide is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, or an isolated polypeptide comprising an immunogenic portion of a polypeptide encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164.

3. The polypeptide of claim 1 or 2, wherein the polypeptide is fused to a second polypeptide to form a fusion protein.

4. The fusion protein of claim 3, wherein the two polypeptides are heterologous.

5. The fusion protein of claim 3, wherein the polypeptides are *Mycobacterium tuberculosis* polypeptides.

6. The fusion protein of claim 3, wherein the second polypeptide is a known *Mycobacterium* antigen.

7. A polynucleotide comprising a nucleotide sequence encoding a fusion protein according to claim 3.

8. A pharmaceutical composition comprising a fusion protein according to claim 3 and a physiologically acceptable carrier.

9. An isolated polynucleotide that specifically hybridizes under moderately stringent conditions to a second polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164.

10. An isolated polynucleotide that specifically hybridizes under highly stringent conditions to a second polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164.

11. An expression vector comprising a polynucleotide according to claim 9 or 10.

12. A host cell transformed with an expression vector according to claim 11.

13. The host cell of claim 12, wherein the host cell is selected from the group consisting of *E. coli*, yeast, and mammalian cells.

14. A method for detecting *Mycobacterium* infection in a biological sample, the method comprising the steps of:

(a) contacting a biological sample with at least one polypeptide according to claim 1 or 2; and

(b) detecting in the sample the presence of antibodies that bind to the polypeptide, thereby detecting *Mycobacterium* infection in the biological sample.

15. The method of claim 14, wherein the polypeptide is bound to a solid support.

16. The method of claim 15, wherein the solid support comprises nitrocellulose, latex or a plastic material.

17. The method of claim 14, wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

18. The method of claim 17, wherein the biological sample is whole blood or serum.

19. The method of claim 14, wherein the *Mycobacterium* infection is a *Mycobacterium tuberculosis* infection.

20. A method for detecting *Mycobacterium* infection in a biological sample, the method comprising the steps of:

(a) contacting the sample with at least two oligonucleotide primers, wherein at least one of the oligonucleotide primers specifically hybridizes under stringent conditions to a polynucleotide according to claim 9; and

(b) detecting in the sample a polynucleotide sequence that is amplified in the presence of the oligonucleotide primers, thereby detecting *Mycobacterium* infection.

21. The method of claim 20, wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. The method of claim 20, wherein the *Mycobacterium* infection is a *Mycobacterium tuberculosis* infection.

23. A method for detecting *Mycobacterium* infection in a biological sample, the method comprising the steps of:

(a) contacting the sample with one or more polynucleotide probes that specifically hybridize to a polynucleotide according to claim 9; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *Mycobacterium* infection.

24. The method of claim 23, wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

25. The method of claim 23, wherein the *Mycobacterium* infection is a *Mycobacterium tuberculosis* infection.

26. A method for detecting *Mycobacterium* infection in a biological sample, the method comprising the steps of:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to claim 1 or 2; and

(b) detecting in the sample a polypeptide that binds to the binding agent, thereby detecting *Mycobacterium* infection in the biological sample.

27. The method of claim 26, wherein the binding agent is a monoclonal antibody.

28. The method of claim 26, wherein the binding agent is a polyclonal antibody.

29. The method of claim 26, wherein the *Mycobacterium* infection is a *Mycobacterium tuberculosis* infection.

30. A diagnostic kit comprising:

(a) one or more polypeptides according to claim 1 or 2; and

(b) a detection reagent.

31. The kit of claim 30, wherein the polypeptide is immobilized on a solid support.

32. The kit of claim 30, wherein the detection reagent comprises a reporter group conjugated to a binding agent.

33. The kit of claim 32, wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.

34. The kit of claim 32, wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.

35. A diagnostic kit comprising at least two oligonucleotide primers, wherein at least one of the oligonucleotide primers specifically hybridizes under stringent conditions to a polynucleotide according to claim 9.

36. A diagnostic kit comprising at least one polynucleotide probe, wherein the polynucleotide probe specifically hybridizes under stringent conditions to a polynucleotide according to claim 9.

37. An antibody that binds to a polypeptide according to claim 1 or 2.

38. The antibody of claim 37, wherein the antibody is a monoclonal antibody.

39. A pharmaceutical composition comprising at least one polypeptide according to claim 1 or 2, and a physiologically acceptable carrier.

40. A pharmaceutical composition comprising a polynucleotide according to claim 9 and a physiologically acceptable carrier.

41. The pharmaceutical composition of claim 39 or 40, wherein the pharmaceutical composition is a vaccine and a non-specific immune response enhancer.

42. The vaccine of claim 41, further comprising a non-specific immune response enhancer.

43. The vaccine of claim 42, wherein the non-specific immune enhancer is an adjuvant.

44. The vaccine of claim 43, wherein the adjuvant is selected from the group consisting of SBAS-2, QS-21, 3D-MPL, GM-CSF, SAF, ISCOMS, MF-59 and RC-529.

45. A method for eliciting or enhancing an immune response to *Mycobacterium* in a patient, the method comprising the step of administering to a patient a pharmaceutical composition according to claims 39 or 40 in an amount effective to elicit or enhance the immune response.

46. A method for inhibiting the development of a *Mycobacterium* infection in a patient, the method comprising the step of administering to a patient an effective amount

of a pharmaceutical composition according to claims 39 or 40, and thereby inhibiting the development of a *Mycobacterium* infection in the patient.

47. A method for inhibiting the development of a *Mycobacterium* infection in a patient, the method comprising the step of administering to a patient an effective amount of an antibody according to claim 37, and thereby inhibiting the development of a *Mycobacterium* infection in the patient.

48. The method of claims 46 or 47, wherein the *Mycobacterium* infection is a *M. tuberculosis* infection.

49. A method for detecting tuberculosis in a patient, the method comprising the steps of:

(a) contacting dermal cells of a patient with at least one polypeptide according to claim 1 or 2; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

50. The method of claim 49, wherein the immune response is induration.

51. A diagnostic kit comprising:

(a) a polypeptide according to claim 1 or 2; and

(b) an apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

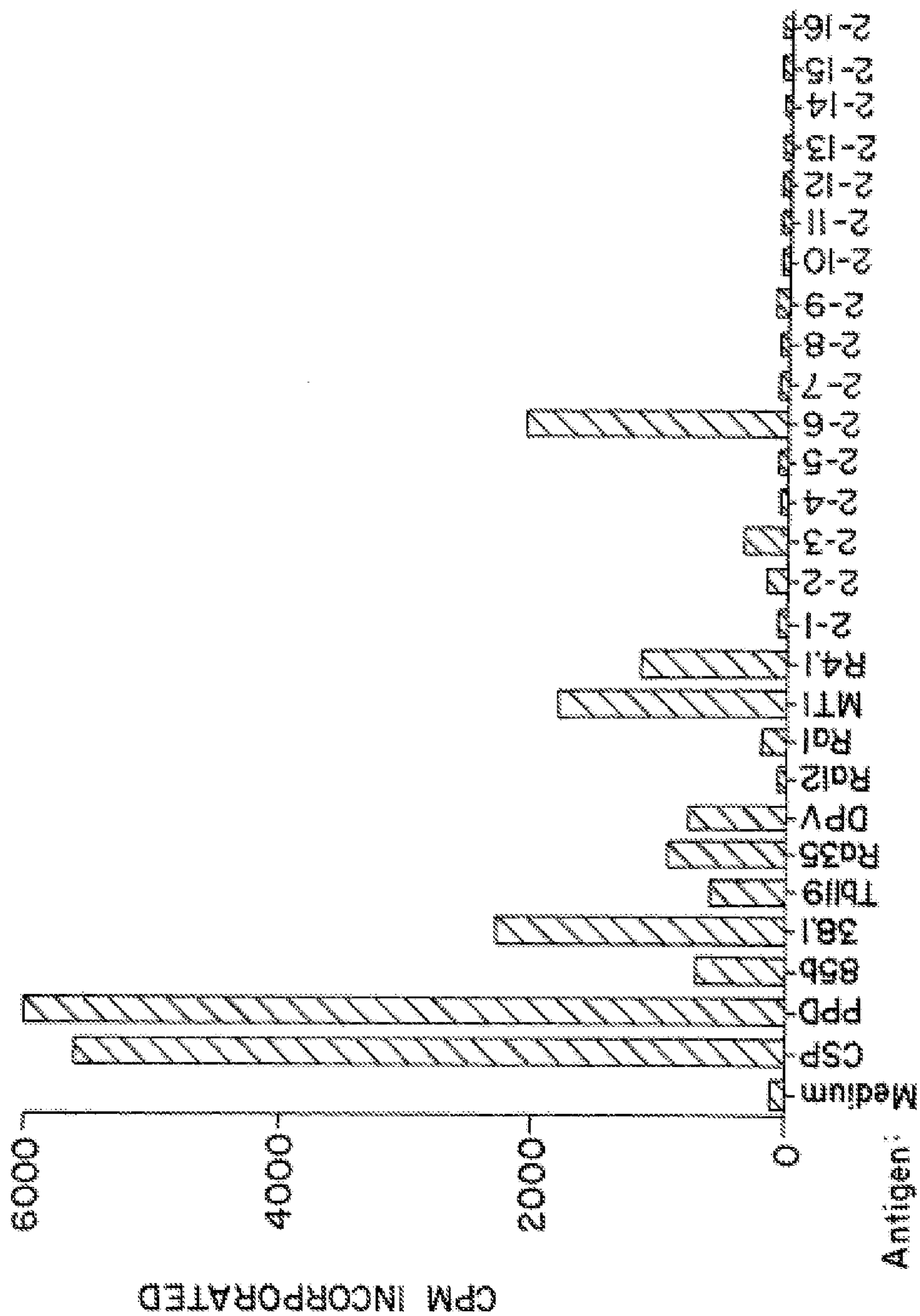
CPM INCORPORATED
6000
4000
2000
0
Antigen:
Medium
CSP
PPD
85b
38.1
TbII9
Ra35
DPV
Ra12
Ra1
MTI
R4.1
2-1
2-2
2-3
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
2-13
2-14
2-15
2-16

FIG. 1A.

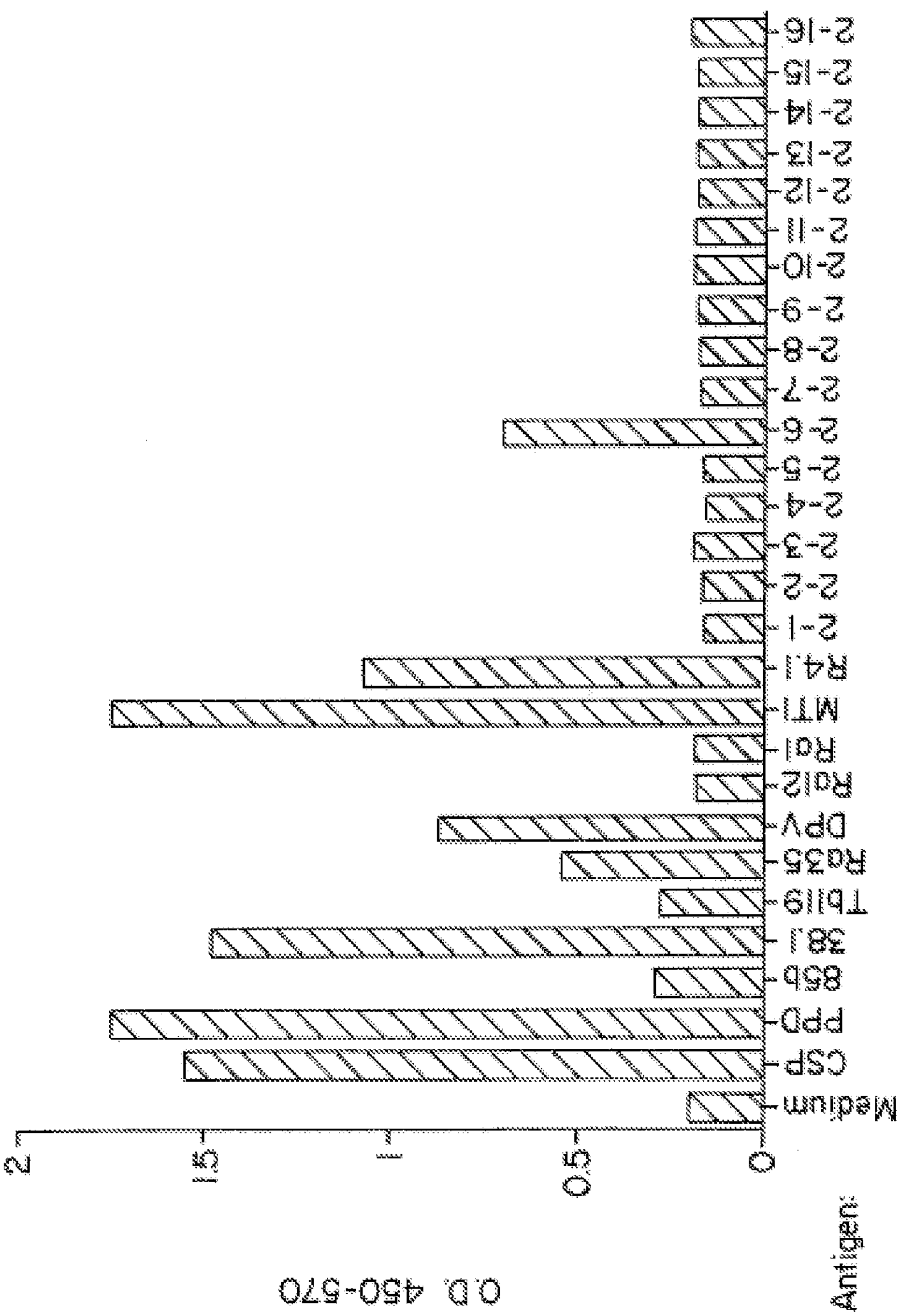
O.D. 450-570
0
0.5
1
1.5
2
Antigen:
Medium
CSP
PPD
85b
38.1
Tbll9
Ra35
DPV
Ral2
Ral
MTI
R4.1
2-1
2-2
2-3
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
2-13
2-14
2-15
2-16

FIG. 1B.

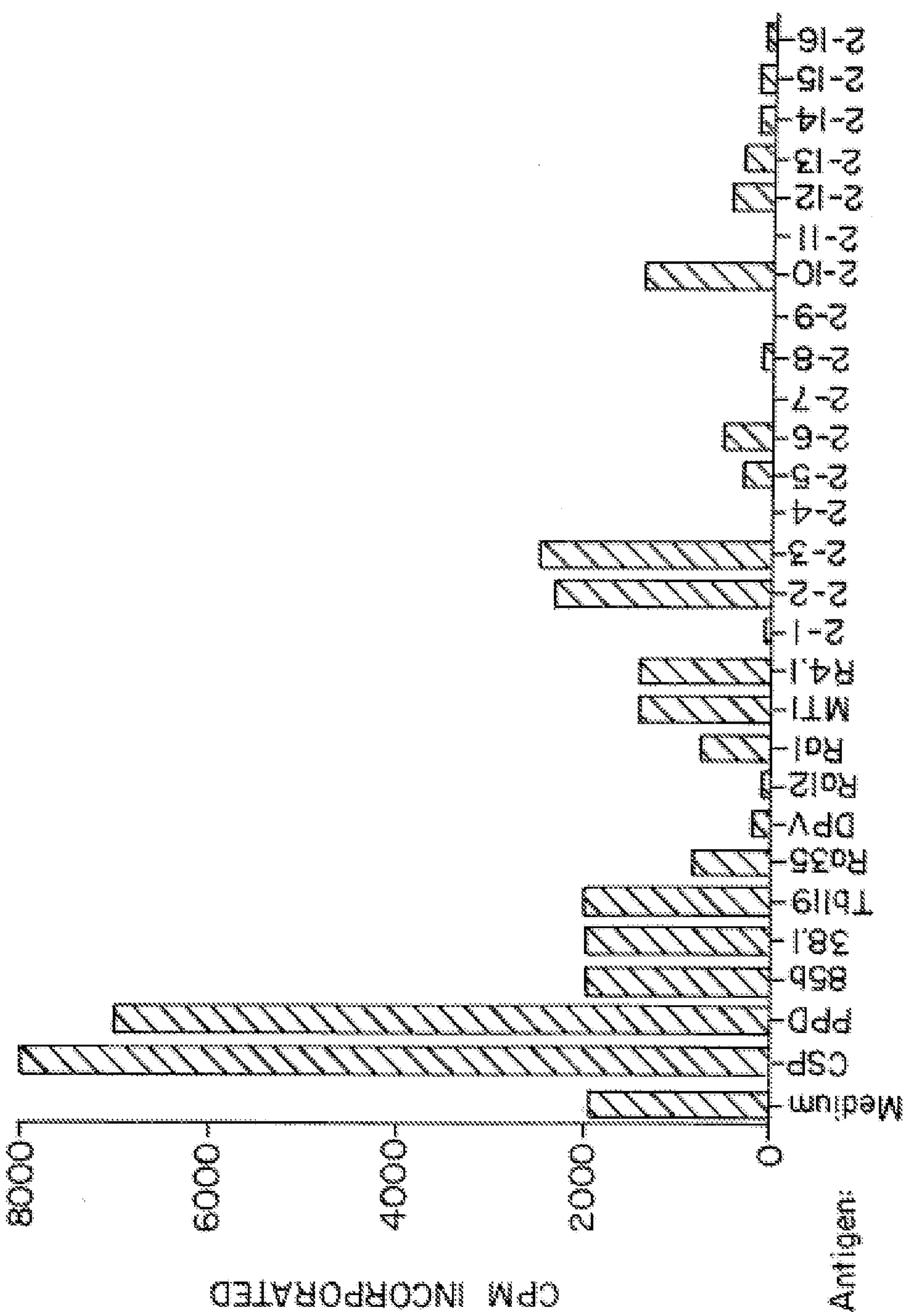
CPM INCORPORATED
0
2000
4000
6000
8000
Antigen:
Medium
CSP
PPD
85b
38.1
TbII9
Ra35
DPV
Ral2
Ral
MTI
R4.1
2-1
2-2
2-3
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
2-13
2-14
2-15
2-16

FIG. 2A.

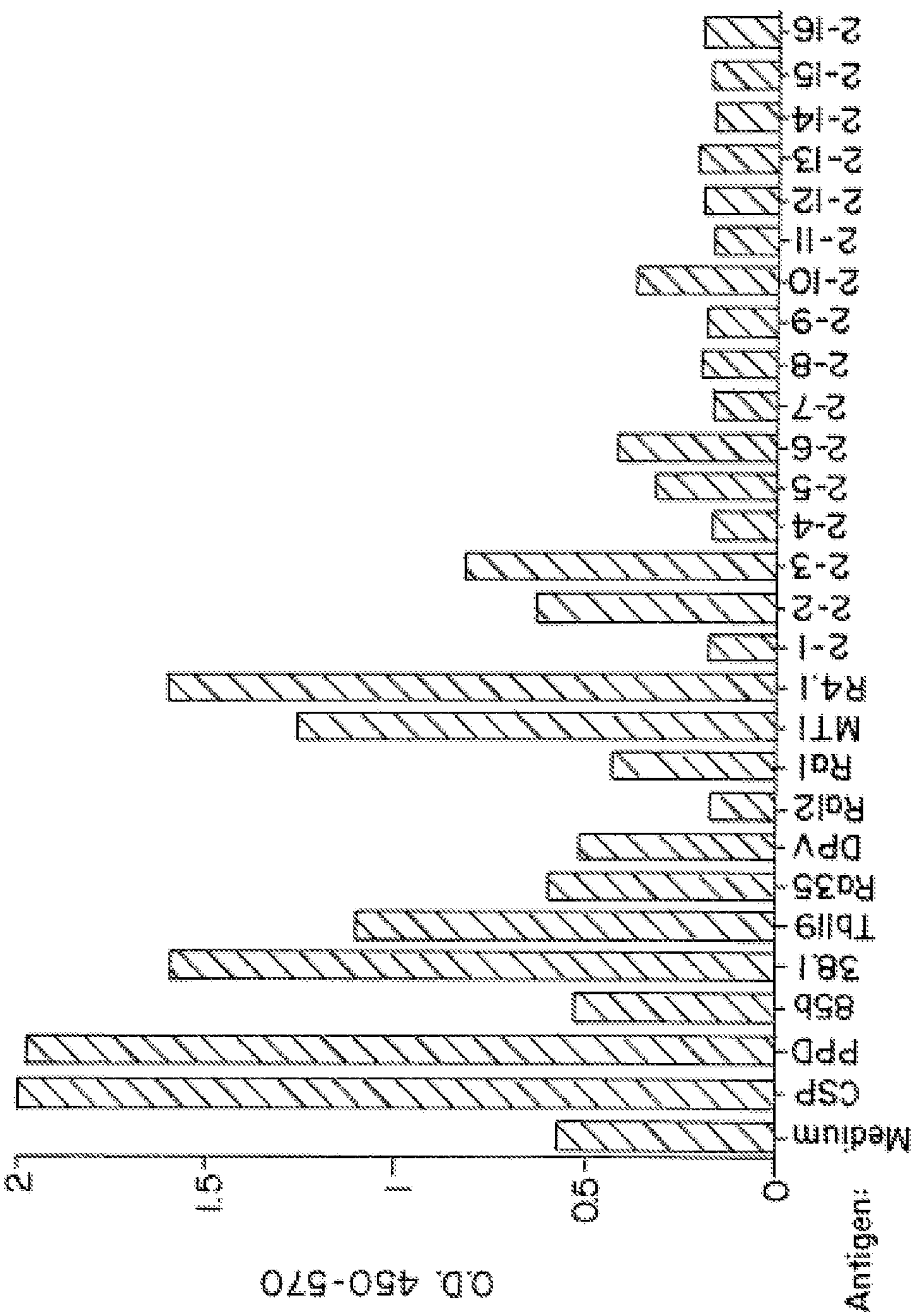
O.D. 450-570
0
0.5
1
1.5
2
Antigen:
Medium
CSP
PPD
85b
38.1
Tbll9
Ra35
DPV
Ra12
Ral
MTI
R4.1
2-1
2-2
2-3
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
2-13
2-14
2-15
2-16

FIG. 2B.

>mTCC#3His.seq
ATGCATCACCATCACCATCACATGAATTATTCGGTGTTGCCGCCGGAGATTAATTCGTTGCGGATG
TTTACCGGTGCGGG
GTCTGCGCCGATGCTTGCGGCATCGGTGGCTTGGGATGGTTTGGCCGCGGAGTTGGCGGTGGCGGC
GTCCTCGTTTGGGT
CGGTGACTTCGGGGTTGGCGGGTCAGTCCTGGCAGGGTGCGGCGGCGGCGGCGATGGCCGCGGCGG
CGGCGCCGTATGCG
GGGTGGTTGGCTGCTGCGGCGGCGCGGGCCGCTGGCGCGTCGGCTCAGGCCAAGGCGGTGGCCAGT
GCGTTTGAGGCGGC
GCGGGCGGCGACGGTGCATCCGATGCTGGTGGCGGCCAACCGTAATGCGTTTGTGCAGTTGGTGTT
GTCGAATCTGTTTG
GGCAGAATGCGCCGGCGATCGCGGCCGCTGAGGCGATGTATGAACAGATGTGGGCCGCCGATGTGG
CCGCGATGGTGGGC
TATCACGGCGGGGCATCGGCGGCCGCGGCGCAGCTGTCGTCGTGGTCAATTGGTCTGCAGCAGGCG
TTGCCAGCTGCGCC
ATCGGCGCTGGCCGCCGCGATCGGCCTCGGCAACATCGGCGTCGGGAACCTGGGCGGCGGGAACAC
CGGTGACTACAATC
TGGGCAGCGGAAATTCCGGCAACGCCAACGTAGGTAGCGGAAACTCCGGCAACGCCAATGTGGGCA
GCGGAAATGACGGT
GCCACGAATTTGGGCAGCGGAAATATCGGCAACACCAATCTCGGCAGCGGAAACGTTGGCAATGTC
AATCTGGGCAGCGG
AAACCGAGGCTTTGGAAACCTCGGCAACGGAAACTTTGGCAGTGGAACCTGGGCAGTGGAAACAC
CGGAAGTACCAACT
TCGGCGGCGGAAATCTCGGTTCCTTCAACTTGGGCAGTGGAAACATCGGCTCCTCCAACATCGGTT
TCGGAAACAACGGC
GACAATAACCTCGGCCTCGGGAACAATGGCAACAACAACATCGGTTTTGGGCTCACCGGCGACAAC
TTGGTGGGCATTGG
CGCGCTGAACTCGGGCATCGGGAATCTAGGTTTCGGGAACTCGGGTAACAACAACATCGGTTTCTT
CAACTCTGGCAACA
ACAACGTGGGCTTCTTCAATTCGGGCAACAACAACTTCGGCTTTGGAAACGCGGGCGACATCAACA
CGGGCTTCGGAAAC
GCCGGCGACACCAACACGGGCTTCGGAAACGCCGGCTTCTTCAATATGGGCATCGGGAACGCGGGC
AACGAAGACATGGG
CGTCGGGAACGGCGGTTCCTTTAACGTGGGCGTTGGCAATGCGGGCAACCAAAGTGTGGGCTTTGG
CAACGCGGGCACCC
TAAACGTGGGCTTCGCAAACGCGGGCAGTATCAATACGGGATTCGCGAACTCGGGCAGCATCAATA
CGGGCGGTTTCGAC
TCGGGCGACCGGAACACCGGGTTTGGAAGCTCGGTCGACCAATCCGTTTCGAGCTCGGGCTTCGGC
AACACCGGCATGAA
TTCCTCAGGCTTCTTTAACACGGGCAATGTTTCGGCTGGCTATGGGAACAACGGTGACGTTCAGTC
GGGCATCAATAACA
CCAACTCCGGCGGCTTCAACGTCGGCTTCTATAACTCGGGTGCCGGCACCGTGGGCATCGCAAACT
CTGGCCTGCAGACC
ACAGGCATTGCGAACTCGGGCACCCTCAACACGGGTGTGGCGAACACGGGTGACCACAGCTCGGGG
GGCTTCAATCAGGG
CAGTGACCAGTCGGGCTTCTTCGGTCAGCCCTAA

FIG. 3.

>mTCC#3-His.pro
MHHHHHHMNYSVLPPEINSLRMFTGAGSAPMLAASVAWDGLAAELAVAASSFGSVTSGLAGQSWQG
AAAAAMAAAAAPYA
GWLAAAAARAAGASAQAKAVASAFEAARAATVHPMLVAANRNAFVQLVLSNLFGQNAPAIAAAEAM
YEQMWAADVAAMVG
YHGGASAAAAQLSSNSIGLQQALPAAPSALAAAIGLGNIGVGNLGGGNTGDYNLGSGNSGNANVGS
GNSGNANVGSGNDG
ATNLGSGNIGNTNLGSGNVGNVNLGSGNRGFGNLGNGNFGSGNLGSGNTGSTNFGGGNLGSFNLGS
GNIGSSNIGFGNNG
DNNLGLGNNGNNNIGFGLTGDNLVGIGALNSGIGNLGFGNSGNNNIGFFNSGNNNVGFFNSGNNNF
GFGNAGDINTGFGN
AGDTNTGFGNAGFFNMGIGNAGNEDMGVGNGGSFNVGVGNAGNQSVGFGNAGTLNVGFANAGSINT
GFANSGSINTGGFD
SGDRNTGFGSSVDQSVSSSGFGNTGMNSSGFFNTGNVSAGYGNNGDVQSGINNTNSGGFNVGFYNS
GAGTVGIANSGLQT
TGIANSGTLNTGVANTGDHSSGGFNQGSDQSGFFGQP.

FIG. 4.

ggatccgaat tctgcacgag ggkygacgac gamctttgca cacgagcgat 50
ggcaaccctc acgtccgcgc aaacccgcg cgaggccgta gagcaattcg 100
tcgagctgat ggtcgacgat ccggtgcgcg ggcgcgtgct gttgctggcg 150
ccggcggtag aaccggccct gacccggtcg ggcgcggagt ggatgcccaa 200
cttcatcgag ttgctgcaac gcaagttgtc ccgcatcgtt gatccagttc 250
tgcagaaact ggtcgccacc agcttgatcg gcgctcttac cggtctgttc 300
accgcatatc tgaacggacg gctgggagcc acccgcaagc aattcatcga 350
ctactgcgtc aacatgttgc tcagcaccgc cgcacctacg caccgcaccg 400
cgagcgggga gaatccgaac a 421

FIG. 5.

gatccgaatt cggcacgagt cgaggccacc gcttccatgg ccaggccacg 50
atyttgatcg gcgtggtggc cacgcccggt gtgaagtgct gttggccgtg 100
atgtcggatt acagtctcgg cgtgcccgac gagacaggcc ttggtgctga 150
cgcggcgcgc gcgtgaagtg gcgctgacac agcacattgg ggtatccgcg 200
gagaccgatc gggccgtcgt ccccaagctg cgccaggcct atgacagcct 250
ggtgtgcggt cgccgccggc ttggcgccat tggagccgag atcgagaacg 300
cggtggccca tcagcgcgcg ctgggccttg acacccggc cggtgcccgt 350
aacttctccc ggtttctcgc caccaaagca cacgacatca cgcgagtgct 400
ggcagcaacc gccgcggaat cccaggccgg cgcggcgcgg ttgcgatccc 450
tggcttcgtc ctatcaggct gtgggatttg gccccaaacc ccaggagccg 500
cctccggatc cagtgccatt tccgccctac cagccgaagg tgtgggcggc 550
gtgccgggcg cgtggccaag accccggacaa ggtcgtcagg acgttccatc 600
acgcgccgat gagcgcgaga ttccgctcgc ttactcgtgc cgaattsgga 650
tctgatatcg ccatggcctt gtcgt 675

FIG. 6.

```
tgatcggtca atgcgcagta ctggtgacct agcgccgccg cggtggccat  50
catctcctcg atcggcgcgg acccgtccga ccagttcgaa tgcagatgca 100
gatccccgcg caatgcggca cggatcgccc ctccaccgag atcctcagcg 150
tcagcgcgta attcagccag caggtccggc tcgcggccag accaggcctg 200
ggcgatgact ttcgcggttt tgggaccgat acccgccagc gactgccagc 250
tgttggcctg gccgtgccgc tgccgc 276
```

FIG. 7.

```
ggatccgaat tctgcacgag gangaagtca tactgccgtc atacacnttt  50
gtctytaccg ccaacgcctt cgtgttgcgc ggtggtgtgc cagtctttgt 100
cgataggcgg cccgacacgc tcaacattga tgaaactcgc atcgtagacg 150
ccatcacccc gcgaaccaag gccatcgtcc ccgttcacta tgccggcgtg 200
gcctgcgaga tggacgcgat catgaagatc gccacgcacc acaacctggc 250
ggtggtcgaa gacgcggccc aaggcgcgat ggcgtcgtat cgtgggcggg 300
cgctcggcag catcggcgac ctgggagcgc tctcatttca cgagaccaag 350
aatgtgattt ccggcgaagg cggcgccctg cttgtcaact cataagactt 400
cctgctccgg gcagakattc tcagggaaaa gggcaccaat mrcagccngc 450
ttcctt                                                 456
```

FIG. 8.

```
gatatcggat cggaattcgg cacgaggtgc ccntgggggg acaactggtg  50
cacaagaggt tcgtccgtcc cggtcctntc gtatagggac aggtttcctc 100
aagtttctga cgcgcgcggc ggatagagac cgaactgtct cacgacgttc 150
taaacccagc tcgcgtgccg ctttaatggg cgaacagccc aacccttggg 200
acctgctcca gccccaggat gcgacgagcc gacatcgagg tgccaaacca 250
tcccgtcgat atggactctt ggggaagatc agcctgttat ccccggggta 300
ccttttatcc gttgagcgac acccccttcca ctcggggggtg c       341
```

FIG. 9.

```
gatccgaatt cagagcggcg acccgtgctc caagctcctt cagcgtcgtc  50
acgggctcat cctatccggc agatcagcag gcggttcctc cgcaaagtgc 100
ggctgcaacc taccgacttc gtgcgcggcg aggaacgcgc cccctggggg 150
tatccgcccg cgtcagacaa cagtgcctcg gtctgatcgg taataggcga 200
ccgcctcgag gtccacatcc gccacctgct cgaaacgtca ggtcttgggg 250
tgcggggtgt accggacggt atgcgcccag atcgtgccgt ctcggaatac 300
gaaagtatcg actccgtcgt cgactcggct gaccgcggaa ttcgcggtcc 350
actccaggaa cagtatgtcg ccctcgaaga tttgggtctt taagtc     396
```

FIG. 10.

```
ggatccgaat tcggcacgag gagtatcagc agaggtcgga gaaggtgctg  50
accgaataca acaacaaggc agccctggaa ccggtaaacc cgccgaagcc 100
tccccccgcc atcaagatcg acccgccccc gcctccgcaa gagcagggat 150
tgatccctgg cttcctgatg cc                               172
```

FIG. 11.

```
ggatccgaat tcggcacgag ccagaacctc gcckgccccg ggcggcagng  50
acaccaactg gscaccacgc cgcggatcgg cmgagcagcg cc          92
```

FIG. 12.

```
gatccgaatt cggcacgaga agaatntgac ccnncnccng tggctgatgc  50
gagagcttnc ttntttcttc cccccantgg ttggacgggg tcgtcacagc 100
gggcattcta agtcccgcgg gccacaaaag gcagtgccgc ggaacttctt 150
ggcccaaacg ggcacccggc tacgtgcgca ccgcgaccgt cgacaactgg 200
tcggcgagcc ggtccgggga atccaccatc gagaacgtcc gtgctccctc 250
gattacctcg aaacgggcgc gcgggatggt cgcggcgagc cgttgaccgt 300
tctcgagtgc gaagaacacg tcatccgccg accacgcgat gagcgccggc 350
ttgtcgaatt caggcagccg ggcggcgact gcggtggtga cttcggtgcg 400
cagcgatagc gagagctgac gcaggtcttc ggcgatggcc gggttggata 450
gcgccggacg aacccaggcc cgggtgagat ggtcgatgtt gtggtgcgac 500
aaaccggcat acgcgcggtt tacgcgcggc cggtgcccgc atcacctgga 550
tcgcggcccg gaacagggtg gccgatttcg cggncaggat cacctgnttt 600
gaggatcgg                                              609
```

FIG. 13.

```
ggatccgaat tcggcacgag tgcggtgcct atctgcgttg gccagtacct  50
cgcggacctg gcgagtgcgg acgcgcaggc tatcgaagtg ggcctaaaga 100
cggcggacgt ggcgcccgtt gccgtacgac ctgcagcggc gccgccgttg 150
cgtgagtctg ccgcggtgcg accggaggcc aggctggtgt cggcggtggc 200
gccagctccc gcgggcacgt cggcgtcggt gctggcttcg gatcggggtg 250
ccggcgtgtt ggggtttgcc gggaccgctg gcaaggantc cnttgggcgt 300
c                                                      301
```

FIG. 14.

```
ggctgctgcg cgcactcgcg ggtctgctgg acgagtggac gccggtgatc  50
gccggcgccg aactgggcga gcacccctac acgccgatca cgccggagtc 100
gatccggcgg gccgcgcagc tcggcgacga cctaccggtg gcgtggaagc 150
accgcagcga gcgctacacc gagaagctgg ccaccccga  caccagcgtc 200
gccgacctgg tcggcgacgt cgacccgatc aaggttgccg agggccgcag 250
cctcggggat c                                           261
```

FIG. 15.

PREDICTED PROTEIN SEQUENCE (SEQ ID NO: 161)

VRHHEGHVAADDDQPQCASFGALTGVIEDIAENQRNAHHQKWRHGRCVEEVHLPVDVGEPRQPTGA
VADQDHRITPVPAHKHTPFRVCQDWHRQPPHRGRADQHLGLDARLCAAACNVLLVDGVQHRPQRHG
PGPRFGFPRVVVACGIRQARVEVERFGGVVPERAHGVGQRNNRVATDRLTDRMPIDRGLGREPRSV
GGQIDRERDQPQRIPAGKHVTPHCPQPRALHLVLTSRRHVERQRHRAEEQHEVHAGPLGGASQSQQ
HPGAEPPPAHTHPRSPHGGGAAAGQQSDVHPPANLIAVDDERAERRDDEERQEAVQQRGPRGDEAD
PVADQQHPGDGADQCRPADPPHDPHHQRHQDHTQQGAGEPPAESVVTEDGLPDRDQLLTDRRVNHQ
AVPGVVFHPMVVQHLPGLGCVMLLVEDGGAGIGQRAQVQEPGHRGQQRDQAGHDPAA

NECLEOTIDE SEQUENCE (SEQ ID NO: 160)

TGAGATTGGCAGACCGGTGAGCACCGGATACAGCCACGCAAAGTTCGTCACCACGAGGGCCACGTA
GCAGCAGACGACGATCAGCCCCAGTGTGCGTCGTTCGGAGCCCTGACCGGGGTGATAGAGGATATC
GCCGAGAACCAGCGAAATGCCCATCACCAGAAATGGCGCCATGGTCGCTGCGTAGAAGAAGTACAT
CTGCCGGTCGATGTCGGCGAACCACGGCAGCCAACCGGCGCAGTAGCCGACCAGGACCACCGCATA
ACGCCAGTCCCGGCGCACAAACATACGCCACCCCGCGTATGCCAGGACTGGCACCGCCAGCCACCA
CATCGCGGGCGTGCCGACCAGCATCTCGGCCTTGACGCACGACTGTGCGCCGCAGCCTGCAACGTC
TTGCTGGTCGATGGCGTACAGCACCGGCCGCAACGACATGGGCCAGGTCCACGGTTTGGATTCCCA
AGGGTGGTAGTTGCCTGCGGAATTCGTCAGGCCCGCGTGGAAGTGGAACGCTTTGGCGGTGTAGTG
CCAGAGCGAGCGCACGGCGTCGGGCAGCGGAACAACCGAGTTGCGACCGACCGCTTGACCGACCGC
ATGCCGATCGATCGCGGTCTCGGACGCGAACCACGGAGCGTAGGTGGCCAGATAGACCGCGAACGG
GATCAACCCCAGCGCATACCCGCTGGGAAGCACGTCACGCCGCACTGTCCCCAGCCACGGTCTTTG
CACTTGGTACTGACGTCGCGCCGCCACGTCGAACGCCAGCGCCATCGCGCCGAAGAACAGCACGAA
GTACACGCCGGACCACTTGGTGGCGCAAGCCAATCCCAGCAGCACCCCGGCGCCGAACCGCCACCA
GCGCACACCCACCCGCGGTCCCCACACGGTGGCGGCGCTGCGGCCGGCCAGCAGAGCGATGTGCAT
CCGTTCGCGAACCTGATCGCGGTCGACGATGAGCGCGCCGAACGCCGCGACGACGAAGAACGTCAG
GAAGCCGTCCAGCAGCGCGGTCCGCGCGGTGACGAAGCTGACCCCGTCGCAGATCAGCAGCACCCC
GGCGATGGCGCCGACCAATGTCGACCGGCTGATCCGCCGCACGATCCGCACCACCAGCGCCACCAG
GACCACACCCAGCAGGGCGCCGGTGAACCGCCAGCCGAATCCGTTGTAACCGAAGATGGCCTCCCC
GATCGCGATCAGCTGCTTACCGACCGGCGGGTGAACCACCAGGCCGTACCCGGGGTTGTCTTCCAC
CCCATGGTTGTTCAGCACCTGCCAGGCCTGGGGTGCGTAATGCTTCTCGTCGAAGATGGGGGTGCC
GGCATCGGTCAGCGAGCCCAGGTTCAGGAACCGGGTCACCGTGGCCAGCAGCGTGATCAGGCCGGT
CACGATCCAGCCGCGTAA

NOTES: UNKNOWN PROTEIN FROM COSMID MTC1237

FIG. 16.

MO-2
PREDICTED PROTEIN SEQUENCE (SEQ ID NO: 163)

VALVVQKYGGSSVADAERIRRVAERIVATKKQGNDVVVVVSAMGDTTDDLLDLAQQVCPAPPPREL
DMLLTAGERISNALVAMAIESLGAHARSFTGSQAGVITTGTHGNAKIIDVTPGRLQTALEEGRVVL
VAGFQGVSQDTKDVTTLGRGGSDTTAVAMAAALGADVCEIYTDVDGIFSADPRIVRNARKLDTVTF
EEMLEMAACGAKVLMLRCVEYARRHNIPVHVRSSYSDRPGTVVVGSIKDVPMEDPILTGVAHDRSE
AKVTIVGLPDIPGYAAKVFRAVADADVNIDMVLQNVSKVEKGKTDITFTCSRDVGPAAVEKLDSLR
NEIGFSQLLYDDHIGKVSLIGAGMRSHPGVTATFCEALAAVGVNIELISTSEIRISVLCRDTELDK
AVVALHEAFGLGGDEEATVYAGTGR

NUCLEOTIDE SEQUENCE (SEQ ID NO. 162)

GTGGCGCTCGTCGTGCAGAAGTACGGCGGATCCTCGGTGGCCGACGCCGAACGGATTCGCCGCGTC
GCCGAACGCATCGTCGCCACCAAGAAGCAAGGCAATGACGTCGTCGTCGTCGTCTGCCATGGGGGA
TACCACCGACGACCTGCTGGATCTGGCTCAGCAGGTGTGCCCGGCGCCGCCGCCTCGGGAGCTGGA
CATGCTGCTTACCGCCGGTGAACGCATCTCGAATGCGTTGGTGGCCATGGCCATCGAGTCGCTCGG
CGCGCATGCCCGGTCGTTCACCGGTTCGCAGGCCGGGGTGATCACCACCGGCACCCACGGCAACGC
CAAGATCATCGACGTCACGCCGGGGCGGCTGCAAACCGCCCTTGAGGAGGGGCGGGTCGTTTTGGT
GGCCGGATTCCAAGGGGTCAGCCAGGACACCAAGGATGTCACGACGTTGGGCCGCGGCGGCTCGGA
CACCACCGCCGTCGCCATGGCCGCCGCGCTGGGTGCCGATGTCTGTGAGATCTACACCGACGTGGA
CGGCATCTTCAGCGCCGACCCGCGCATCGTGCGCAACGCCCGAAAGCTCGACACCGTGACCTTCGA
GGAAATGCTCGAGATGGCGGCCTGCGGCGCCAAGGTGCTGATGCTGCGCTGCGTGGAATACGCTCG
CCGCCATAATATTCCGGTGCACGTCCGGTCGTCGTACTCGGACAGACCGGGCACCGTCGTTGTCGG
ATCGATCAAGGACGTACCCATGGAAGACCCCATCCTGACCGGAGTCGCGCACGACCGCAGCGAGGC
CAAGGTGACCATCGTCGGGCTGCCCGACATCCCCGGGTATGCGGCCAAGGTGTTTAGGGCGGTGGC
CGACGCCGACGTCAACATCGACATGGTGCTGCAGAACGTCTCCAAGGTCGAGGACGGCAAGACCGA
CATCACCTTCACCTGCTCCCGCGACGTCGGGCCCGCCGCCGTGGAAAAACTGGACTCGCTCAGAAA
CGAGATCGGCTTCTCACAGCTGCTGTACGACGACCACATCGGCAAGGTATCGCTGATCGGTGCCGG
CATGCGCAGCCACCCCGGGGTCACCGCGACGTTCTGTGAGGCGCTGGCGGCGGTGGGGGTCAACAT
CGAGCTGATCTCCACCTCGGAGATCAGGATCTCGGTGTTGTGCCGCGACACCGAACTGGACAAGGC
CGTGGTCGCGCTGCATGAAGCGTTCGGGCTCGGCGGCGACGAGGAGGCCACGGTGTACGCGGGGAC
GGGACGGTAGATGGGCCTGTCAATAGGGATCGTGGGGGCCACCGGTCAGGTGGGTCAGGTCATGCG
CACGTTGCTCGACGAGCGGGATTTCCCGGCGAGCGCGGTGCGGTTCTTCGCGTCGGCCCGATCGCA
GGGCCGCAAGCTGGCCTTCCGCGGCCAGGAGATCGAAGTGGAAGACGCCGAGACGGCCGACCCGAG
CGGGCTGGATATCGCGTTGTTCTCCGCCGGCTCGGCCATGTCGAAGGTGCAGGCGCCCCGCTTTGC
GGCGGCCGGAGTCACGGTGATCGACAACTCGTCGGCGTGGCGTAAGGACCCCGACGTGCCGTTGGT
GGTGTCCGAGGTGAACTTTGAACGCGACGCGCACCGCCGGCCCAAGGCTCGTGCCGCTCGTGCCGA
ATTCGGCACGAGCCGACGTGGTCGGCAACGTCCTGGATCGCGGGCAGCTGGTTGTTGAGGATGAAT
CCGTCCACCAGGTGGTAGGAGCCGAACGAAGATTCCACCGTCGTCGTCAACGTGGCCGCATTGCCG
TACGAATCGACGACGCTGAGGTGGCTGGTGCCATGCTCAGGCACTGGCGGGGCGACGGCCGTCGGT
GCGCCGAAGTCCC

NOTES: M.tb aspartokinase

FIG. 17.

>Full-length TbH4/XP-1 (MTB48) Open Reading Frame (SEQ ID NO: 164)

ATGACGCAGTCGCAGACCGTGACGGTGGATCAGCAAGAGATTTTGAACAGGGCCAACGAGGTGGAG
GCCCCGATGGCGGACCCACCGACTGATGTCCCCATCACACCGTGCGAACTCACGGCGGCTAAAAAC
GCCGCCCAACAGCTGGTATTGTCCGCCGACAACATGCGGGAATACCTGGCGGCCGGTGCCAAAGAG
CGGCAGCGTCTGGCGACCTCGCTGCGCAACGCGGCCAAGGCGTATGGCGAGGTTGATGAGGAGGCT
GCGACCGCGCTGGACAACGACGGCGAAGGAACGTGCAGGCAGAATCGGCCGGGGCCGTCGGAGGGG
ACAGTTCGGCCGAACTAACCGATACGCCGAGGGTGGCCACGGCCGGTGAACCCAACTTCATGGATC
TCAAAGAAGCGGCAAGGAAGCTCGAAACGGGCGACCAAGGCGCATCGCTCGCGCACTTTGCGGATG
GGTGGAACACTTTCAACCTGACGCTGCAAGGCGACGTCAAGCGGTTCCGGGGGTTTGACAACTGGG
AAGGCGATGCGGCTACCGCTTGCGAGGCTTCGCTCGATCAACAACGGCAATGGATACTCCACATGG
CCAAATTGAGCGCTGCGATGGCCAAGCAGGCTCAATATGTCGCGCAGCTGCACGTGTGGGCTAGGC
GGGAACATCCGACTTATGAAGACATAGTCGGGCTCGAACGGCTTTACGCGGAAAACCCTTCGGCCC
GCGACCAAATTCTCCCGGTGTACGCGGAGTATCAGCAGAGGTCGGAGAAGGTGCTGACCGAATACA
ACAACAAGGCAGCCCTGGAACCGGTAAACCCGCCGAAGCCTCCCCCCGCCATCAAGATCGACCCGC
CCCCGCCTCCGCAAGAGCAGGGATTGATCCCTGGCTTCCTGATGCCGCCGTCTGACGGCTCCGGTG
TGACTCCCGGTACCGGGATGCCAGCCGCACCGATGGTTCCGCCTACCGGATCGCCGGGTGGTGGCC
TCCCGGCTGACACGGCGGCGCAGCTGACGTCGGCTGGGCGGGAAGCCGCAGCGCTGTCGGGCGACG
TGGCGGTCAAAGCGGCATCGCTCGGTGGCGGTGGAGGCGGCGGGGTGCCGTCGGCGCCGTTGGGAT
CCGCGATCGGGGGCGCCGAATCGGTGCGGCCCGCTGGCGCTGGTGACATTGCCGGCTTAGGCCAGG
GAAGGGCCGGCGGCGGCGCCGCGCTGGGCGGCGGTGGCATGGGAATGCCGATGGGTGCCGCGCATC
AGGGACAAGGGGGCGCCAAGTCCAAGGGTTCTCAGCAGGAAGACGAGGCGCTCTACACCGAGGATC
GGGCATGGACCGAGGCCGTCATTGGTAACCGTCGGCGCCAGGACAGTAAGGAGTCGAAG

FIG. 18.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Corixa Corporation
(B) STREET: Suite 200, 1124 Columbia Street
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 98104
(G) TELEPHONE: (206) 754-5830
(H) TELEFAX: (206) 754-5994
(I) TELEX:

(ii) TITLE OF INVENTION: Compounds for Immunotherapy and Diagnosis of Tuberculosis and Methods of Their Use

(iii) NUMBER OF SEQUENCES: 144

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Townsend and Townsend and Crew LLP
(B) STREET: Two Embarcadero Center, Eighth Floor
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94111-3834

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM compatible
(C) OPERATING SYSTEM: Windows
(D) SOFTWARE: FastSEQ for Windows Version 2.0b

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US98/10407
(B) FILING DATE: 20-MAY-1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/859,381
(B) FILING DATE: 20-MAY-1997

(A) APPLICATION NUMBER: US 09/073,010
(B) FILING DATE: 05-MAY-1998

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bastian, Kevin L.
(B) REGISTRATION NUMBER: 34,774
(C) REFERENCE/DOCKET NUMBER: 14058-87-1PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 576-0200
(B) TELEFAX: (415) 576-0300
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1886 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGCTCTGGTG ACCACCAACT TCTTCGGTGT CAACACCATC CCGATCGCCC TCAACGAGGC     60
CGACTACCTG CGCATGTGGA TCCAGGCCGC CACCGTCATG AGCCACTATC AAGCCGTCGC    120
GCACGAAATC TGGTGTCTCC ATGAATANGC CAGTTCGGGA AAGCCGTNGG CCAGTATCAC    180
CACNNGTGCG CCGGGCTCAC CNGCCTCGAC CACTCGCAGT CGCACGCCGT TGGTATCAAC    240
TAACCGTNCN GTANGTGCGC CCATCGTCTC ACCAAATCAC ACCGGGCACC GGCCTGAGAA    300
GGGCTTGGGG AGCANCCAGA GGCGATTGTC GCGGGTGCTG CCGCGCATCA TTGATCGGCC    360
GGCCGGACCA NTCGGGCCTC CCTTGACGTC CGGATCNCAC TTCCTGTGCA GCTGGCATNG    420
CTACAGCTCA CAGTGACTGC CCCACGATTG CCNGCCAGGT CCAGTTCAAA TTCCGGTGAA    480
TTCGCNGACA AAAGCAGCAG GTCAACCAAC CGCAGTCAGT CGAGGGTCCC AAACGTGAGC    540
CAATCGGTGA AATGGCTTGC TGCAGTGACA CCNGTCACAG GCTTAGCCGA CAGCACCGGA    600
ATAGCTCAGG CGGGCTATAG AGTCCTATAG AAACATTTGC TGATAGAATT AACCGCTNTC    660
TTGGCGTGAT CTTGATACGG CTCGCCGTGC GACCGGTTGG CTCAGTAGCT GACCACCATG    720
TAACCCATCC TCGGCAGGTG TCTACTAAGG CGAGACACCG CATTGGTGGG GCTGCATCGC    780
AAATCGGTCC GAGCATGTAG CACTGCCGTT ATCCCGGGAT AGCAAACCAC CCGGAACCAG    840
GGCTATCCCA GTCGCTCTCC GACGGAGGCC GTTTCGCTTT CCGTTGCCCG ATAACTCCCG    900
AGTGGATATC GGCGTTATCA NATTCAGGCT TTTCTTCGCA AGGTACCGGT GTTCGCTATA    960
TTCGGATATC TCGGACGGAT AATTACTAAA ACTTCAGTGG TTTAGATAAG GCCGCCGCAA   1020
TACTTCGCCG ATCTTGCCGA GCGCAACGGA TTTCCATCGT CGGTTTTCGT CGCCTTATCA   1080
AACATGATCG GAGATAATGA CAGATCGGCC TAGCTAGGTG TTTAGCGGAC GCGATTTAGG   1140
ACAACCGAGA TTTGCTTTGC CTCGCAACCA TGAGAGCGCC CCGCTTCGAC GCCGAATCGG   1200
GTGAGTGATG GTNGGTTAGC ACAGCCCTGA TTGCGCCACC GGCGAGGTNA TTGTGCCCGC   1260
CACGAGGCCG CCGCCGGCTA GCCCCATGAG CACGNTATAT AGACTCTCCT GCAACAGATC   1320
TCATACCGAT CGAAGGCGAA GCGCAGGCAT CGACGTCGGA GACACTGCCT TGGGATCGCG   1380
CCGCCTACAC GGCGGTTGGC GCATTGTCGC AGCGCAGTTG CAGGAGGGCA AATGTGCGCA   1440
GACGATGTAG TCGACAACAA GTGNACATGC CGTCTTCACG AACTCAAAAC TGACGATCTG   1500
CTTAGCATGA AAAAAACTGT TGACATCGGC CAAGCATGAC AGCCAGACTG TAGGCCTACG   1560
CGTGCAATGC AGAACCAAGG NTATGCATGG AATCGACGAC CGTTGAGATA GGCGGCAGGC   1620
ATGAGCAGAG CGTTCATCAT CGATCCAACG ATCAGTGCCA TTGACGGCTT GTACGACCTT   1680
CTGGGGATTG GAATACCCAA CCAAGGGGGT ATCCTTTACT CCTCACTAGA GTACTTCGAA   1740
AAAGCCCTGG AGGAGCTGGC AGCAGCGTTT CCGGGTGATG GCTGGTTAGG TTCGGCCGCG   1800
GACAAATACG CCGGCAAAAA CCGCAACCAC GTGAATTTTT TCCAGGAACT GGCAGACCTC   1860
GATCGTCAGC TCATCAGCCT GATCCA                                        1886
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
GGCACGCGCT GGCCGCGCAA TACACCGAAA TTGCAACGGA ACTCGCAAGC GTGCTCGCTG     60
```

```
CGGTGCAGGC AAGCTCGTGG CAGGGGCCCA GCGCCGACCG GTTCGTCGTC GCCCATCAAC    120
CGTTCCGGTA TTGGCTAACC CACGCTGCCA CGGTGGCCAC CGCAGCAGCC GCCGCGCACN    180
AAACGGCCGC CGCCGGGTAT ACGTCCGCAT TGGGGGGCAT GCCTACGCTA GCCGAGTTGG    240
CGGCCAACCA TGCCATGCAC GGCGCTCTGG TGACCACCAA CTTCTTCGGT GTCAACACCA    300
TCCCGATCGC CCTCAACGAG GCCGACTACC TGCGCATGTG GATCCAGGCC GCCACCGTCA    360
TGAGCCACTA TCAAGCCGTC GCGCACGAAA GCGTGGCGGC GACCCCCAGC ACGCCGCCGG    420
CGCCGCAGAT AGTGACCAGT GCGGCCAGCT CGGCGGCTAG CAGCAGCTTC CCCGACCCGA    480
CCAAATTGAT CCTGCAGCTA CTCAAGGATT TCCTGGAGCT GCTGCGCTAT CTGGCTGTTG    540
AGCTGCTGCC GGGGCCGCTC GGCGACCTCA TCGCCCAGGT GTTGGACTGG TTCATCTCGT    600
TCGTGTCCGG TCCAGTCTTC ACGTTTCTCG CCTACCTGGT GCTGGACCCA CTGATCTATT    660
TCGGACCGTT CGCCCCGCTG ACGAGTCCGG TCCTGTTGCC TGCTGTGGAG TTACGCAACC    720
GCCTCAAAAC CGCCACCGGA CTGACGCTGC CACCTACCGT GATTTTCGAT CATCCCACTC    780
CCACTGCGGT CGCCGAGTAT GTCGCCCAGC AAATGTCTGG CAGCCGCCCA ACGGAATCCG    840
GTGATCCGAC GTCGCAGGTT GTCGAACCCG CTCGTGCCGA ATTCGGCACG AGTGCTGTTC    900
ATCAAATCCC CCCGAGACCT GCGGACACCC GGCGCGCTTG CCGACATCGA GATGATGTCC    960
CGCGAGATAG CAGAATTGCC CAACATCGTG ATGGTGCGGG GCTTGACCCG ACCGAACGGG   1020
GAACCTCTGA AGGAGACCAA GGTCTCGTTT CAGGCTGGTG AAGTGGGCGG CAAGCTCGAC   1080
GAAGCGACCA CCCTGCTCGA AGAGCACGGA GGCGAGCTGG ACCAGCTGAC CGGCGGTGCG   1140
CACCAGTTGG CCGACGCCCT CGCCCAAATA CGCAACGAAA TCAATGGGGC CGTGGCCAGC   1200
TCGAGCGGGA TAGTCAACAC CCTGCAGGCC ATGATGGACC TGATGGGCGG TGACAAGACC   1260
ATCCGACAAC TGGAAAATGC GTCCCAATAT GTCGGGCGCA TGCGGGCTCT GGGGGACAAT   1320
CTGAGCGGGA CCGTCACCGA TGCCGAACAA ATCGCCACTT GGGCCAGCCC TATGGTCAAC   1380
GCCCTCAACT CCAGCCCGGT GTGTAACAGC GATCCCGCCT GTCGGACGTC GCGCGCACAG   1440
TTGGCGGCGA TTGTCCAGGC GCAGGACGAC GGCCTGCTCA GGTCCATCAG AGCGCTAGCC   1500
GTCACCCTGC AACAGACGCA GGAATACCAG ACACTCGCCC GGACGGTGAG CACACTGGAC   1560
GGGCAACTGA AGCAAGTCGT CAGCACCCTC AAAGCGGTCG ACGGCCTACC CACCAAATTG   1620
GCTCAAATGC AGCAAGGAGC CAACGCTCTC GCCGACGGCA GCGCAGCGCT GGCGGCAGGC   1680
GTGCAGGAAT TGGTCGATCA GGTCAAAAAG ATGGGCTCAG GGCTCAACGA GGCCGCCGAC   1740
TTCCTGTTGG GGATCAAGCG GGATGCGGAC AAGCCGTCAA TGGCGGGCTT CAACATTCCA   1800
CCGCAGATTT TTTCGAGGGA CGAGTTCAAG AAGGGCGCCC AGATTTTCCT GTCGGCCGAT   1860
GGTCATGCGG CGCGGTACTT CGTGCAGAGC GCGCTGAATC CGGCCACCAC CGAGGCGATG   1920
GATCAGGTCA ACGATATCCT CCGTGTTGCG GATTCCGCGC GACCGAATAC CGAACTCGAG   1980
GATGCCACGA TAGGTCTGGC GGGGGTTCCG ACTGCGCTGC GGGATATCCG CGACTACTAC   2040
AACAGCGATA TGAAATTCAT CGTCATTGCG ACGATCGTTA TCGTATTCTT GATTCTCGTC   2100
ATTCTGNTGC GCGCACTTGT GGNTCCGATA TATCTGATAG GCTCGGTGCT GATTTCTTAC   2160
TTGTCGGCCC TAGGCATAGG AACTTTCGTT TTCCAATTGA TACTGGGCCA GGAAATGCAT   2220
TGGAGCCTGC CGGGACTGTC CTTCATATTA TTGGTTGCCA TCGGCGCTGA CTACAACATG   2280
CTGCTCATTT CACGCATCCG CGACG                                         2305
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1742 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CCGCTCTCTT TCAACGTCAT AAGTTCGGTG GGCCAGTCGG CCGCGCGTGC ATATGGCACC     60
AATAACGCGT GTCCCATGGA TACCCGGACC GCACGACGGT AGAGCGGATC AGCGCAGCCG    120
GTGCCGAACA CTACCGCGTC CACGCTCAGC CCTGCCGCGT TGCGGAAGAT CGAGCCCAGG    180
TTCTCATGGT CGTTAACGCC TTCCAACACT GCGACGGTGC GCGCCCCGGC GACCACCTGA    240
GCAACGCTCG GCTCCGGCAC CCGGCGCGCG GCTGCCAACA CCCCACGATT GAGATGGAAG    300
CCGATCACCC GTGCCATGAC ATCAGCCGAC GCTCGATAGT ACGGCGCGCC GACACCGGCC    360
AGATCATCCT TGAGCTCGGC CAGCCGGCGG TCGGTGCCGA ACAGCGCCAG CGGCGTGAAC    420
```

```
CGTGAGGCCA GCATGCGCTG CACCACCAGC ACACCCTCGG CGATCACCAA CGCCTTGCCG     480
GTCGGCAGAT CGGGACNACN GTCGATGCTG TTCAGGTCAC GGAAATCGTC GAGCCGTGGG     540
TCGTCGGGAT CGCAGACGTC CTGAACATCG AGGCCGTCGG GGTGCTGGGC ACAACGGCCT     600
TCGGTCACGG GCTTTCGTCG ACCAGAGCCA GCATCAGATC GGCGGCGCTG CGCAGGATGT     660
CACGCTCGCT GCGGTTCAGC GTCGCGAGCC GCTCAGCCAG CCACTCTTGC AGAGAGCCGT     720
TGCTGGGATT AATTGGGAGA GGAAGACAGC ATGTCGTTCG TGACCACACA GCCGGAAGCC     780
CTGGCAGCTG CGGCGGCGAA CCTACAGGGT ATTGGCACGA CAATGAACGC CCAGAACGCG     840
GCCGCGGCTG CTCCAACCAC CGGAGTAGTG CCCGCAGCCG CCGATGAAGT ATCAGCGCTG     900
ACCGCGGCTC AGTTTGCTGC GCACGCGCAG ATGTACCAAA CGGTCAGCGC CCAGGCCGCG     960
GCCATTCACG AAATGTTCGT GAACACGCTG GTGGCCAGTT CTGGCTCATA CGCGGCCACC    1020
GAGGCGGCCA ACGCAGCCGC TGCCGGCTGA ACGGGCTCGC ACGAACCTGC TGAAGGAGAG    1080
GGGGAACATC CGGAGTTCTC GGGTCAGGGG TTGCGCCAGC GCCCAGCCGA TTCAGNTATC    1140
GGCGTCCATA ACAGCAGACG ATCTAGGCAT TCAGTACTAA GGAGACAGGC AACATGGCCT    1200
CACGTTTTAT GACGGATCCG CATGCGATGC GGGACATGGC GGGCCGTTTT GAGGTGCACG    1260
CCCAGACGGT GGAGGACGAG GCTCGCCGGA TGTGGGCGTC CGCGCAAAAC ATTTCCGGTG    1320
CGGGCTGGAG TGGCATGGCC GAGGCGACCT CGCTAGACAC CATGACCTAG ATGAATCAGG    1380
CGTTTCGCAA CATCGTGAAC ATGCTGCACG GGGTGCGTGA CGGGCTGGTT CGCGACGCCA    1440
ACAANTACGA ACAGCAAGAG CAGGCCTCCC AGCAGATCCT GAGCAGNTAG CGCCGAAAGC    1500
CACAGCTGNG TACGNTTTCT CACATTAGGA GAACACCAAT ATGACGATTA ATTACCAGTT    1560
CGGGGACGTC GACGCTCATG GCGCCATGAT CCGCGCTCAG GCGGCGTCGC TTGAGGCGGA    1620
GCATCAGGCC ATCGTTCGTG ATGTGTTGGC CGCGGGTGAC TTTTGGGGCG GCGCCGGTTC    1680
GGTGGCTTGC CAGGAGTTCA TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAGCA    1740
GG                                                                   1742
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2836 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
GTTGATTCCG TTCGCGGCGC CGCCGAAGAC CACCAACTCC GCTGGGGTGG TGGCACAGGC      60
GGTTGCGTCG GTCAGCTGGC CGAATCCCAA TGATTGGTGG CTCNGTGCGG TTGCTGGGCT     120
CGATTACCCC CACGGAAAGG ACGACGATCG TTCGTTTGCT CGGTCAGTCG TACTTGGCGA     180
CGGGCATGGC GCGGTTTCTT ACCTCGATCG CACAGCAGCT GACCTTCGGC CCAGGGGGCA     240
CAACGGCTGG CTCCGGCGGA GCCTGGTACC CAACGCCACA ATTCGCCGGC CTGGGTGCAG     300
GCCCGGCGGT GTCGGCGAGT TTGGCGGGCG CGGAGCCGGT CGGGAGGTTG TCGGTGCCGC     360
CAAGTTGGGC CGTCGCGGCT CCGGCCTTCG CGGAGAAGCC TGAGGCGGGC ACGCCGATGT     420
CCGTCATCGG CGAAGCGTCC AGCTGCGGTC AGGGAGGCCT GCTTCGAGGC ATACCGCTGG     480
CGAGAGCGGG GCGGCGTACA GGCGCCTTCG CTCACCGATA CGGGTTCCGC CACAGCGTGA     540
TTACCCGGTC TCCGTCGGCG GGATAGCTTT CGATCCGGTC TGCGCGGCCG CCGGAAATGC     600
TGCAGATAGC GATCGACCGC GCCGGTCGGT AAACGCCGCA CACGGCACTA TCAATGCGCA     660
CGGCGGGCGT TGATGCCAAA TTGACCGTCC CGACGGGGCT TTATCTGCGG CAAGATTTCA     720
TCCCCAGCCC GGTCGGTGGG CCGATAAATA CGCTGGTCAG CGCGACTCTT CCGGCTGAAT     780
TCGATGCTCT GGGCGCCCGC TCGACGCCGA GTATCTCGAG TGGGCCGCAA ACCCGGTCAA     840
ACGCTGTTAC TGTGGCGTTA CCACAGGTGA ATTTGCGGTG CCAACTGGTG AACACTTGCG     900
AACGGGTGGC ATCGAAATCA ACTTGTTGCG TTGCAGTGAT CTACTCTCTT GCAGAGAGCC     960
GTTGCTGGGA TTAATTGGGA GAGGAAGACA GCATGTCGTT CGTGACCACA CAGCCGGAAG    1020
CCCTGGCAGC TGCGGCGGCG AACCTACAGG GTATTGGCAC GACAATGAAC GCCCAGAACG    1080
CGGCCGCGGC TGCTCCAACC ACCGGAGTAG TGCCCGCAGC CGCCGATGAA GTATCAGCGC    1140
TGACCGCGGC TCAGTTTGCT GCGCACGCGC AGATGTACCA AACGGTCAGC GCCCAGGCCG    1200
CGGCCATTCA CGAAATGTTC GTGAACACGC TGGTGGCCAG TTCTGGCTCA TACGCGGCCA    1260
CCGAGGCGGC CAACGCAGCC GCTGCCGGCT GAACGGGCTC GCACGAACCT GCTGAAGGAG    1320
```

AGGGGGAACA TCCGGAGTTC TCGGGTCAGG GGTTGCGCCA GCGCCCAGCC GATTCAGCTA 1380
TCGGCGTCCA TAACAGCAGA CGATCTAGGC ATTCAGTACT AAGGAGACAG GCAACATGGC 1440
CTCACGTTTT ATGACGGATC CGCATGCGAT GCGGGACATG GCGGGCCGTT TTGAGGTGCA 1500
CGCCCAGACG GTGGAGGACG AGGCTCGCCG GATGTGGGCG TCCGCGCAAA ACATTTCCGG 1560
TGCGGGCTGG AGTGGCATGG CCGAGGCGAC CTCGCTAGAC ACCATGACCT AGATGAATCA 1620
GGCGTTTCGC AACATCGTGA ACATGCTGCA CGGGGTGCGT GACGGGCTGG TTCGCGACGC 1680
CAACAACTAC GAACAGCAAG AGCAGGCCTC CCAGCAGATC CTGAGCAGCT AGCGCCGAAA 1740
GCCACAGCTG CGTACGCTTT CTCACATTAG GAGAACACCA ATATGACGAT TAATTACCAG 1800
TTCGGGGACG TCGACGCTCA TGGCGCCATG ATCCGCGCTC AGGCGGCGTC GCTTGAGGCG 1860
GAGCATCAGG CCATCGTTCG TGATGTGTTG GCCGCGGGTG ACTTTTGGGG CGGCGCCGGT 1920
TCGGTGGCTT GCCAGGAGTT CATTACCCAG TTGGGCCGTA ACTTCCAGGT GATCTACGAG 1980
CAGGCCAACG CCCACGGGCA GAAGGTGCAG GCTGCCGGCA ACAACATGGC GCAAACCGAC 2040
AGCGCCGTCG GCTCCAGCTG GGCCTAAAAC TGAACTTCAG TCGCGGCAGC ACACCAACCA 2100
GCCGGTGTGC TGCTGTGTCC TGCAGTTAAC TAGCACTCGA CCGCTGAGGT AGCGATGGAT 2160
CAACAGAGTA CCCGCACCGA CATCACCGTC AACGTCGACG GCTTCTGGAT GCTTCAGGCG 2220
CTACTGGATA TCCGCCACGT TGCGCCTGAG TTACGTTGCC GGCCGTACGT CTCCACCGAT 2280
TCCAATGACT GGCTAAACGA GCACCCGGGG ATGGCGGTCA TGCGCGAGCA GGGCATTGTC 2340
GTCAACGACG CGGTCAACGA ACAGGTCGCT GCCCGGATGA AGGTGCTTGC CGCACCTGAT 2400
CTTGAAGTCG TCGCCCTGCT GTCACGCGGC AAGTTGCTGT ACGGGGTCAT AGACGACGAG 2460
AACCAGCCGC CGGGTTCGCG TGACATCCCT GACAATGAGT TCCGGGTGGT GTTGGCCCGG 2520
CGAGGCCAGC ACTGGGTGTC GGCGGTACGG GTTGGCAATG ACATCACCGT CGATGACGTG 2580
ACGGTCTCGG ATAGCGCCTC GATCGCCGCA CTGGTAATGG ACGGTCTGGA GTCGATTCAC 2640
CACGCCGACC CAGCCGCGAT CAACGCGGTC AACGTGCCAA TGGAGGAGAT CTCGTGCCGA 2700
ATTCGGCACG AGGCACGAGG CGGTGTCGGT GACGACGGGA TCGATCACGA TCATCGACCG 2760
GCCGGGATCC TTGGCGATCT CGTTGAGCAC GACCCGGGCC CGCGGGAAGC TCTGCGACAT 2820
CCATGGGTTC TTCCCG 2836

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 900 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AACATGCTGC ACGGGGTGCG TGACGGGCTG GTTCGCGACG CCAACAACTA CGAGCAGCAA 60
GAGCAGGCCT CCCAGCAGAT CCTCAGCAGC TAACGTCAGC CGCTGCAGCA CAATACTTTT 120
ACAAGCGAAG GAGAACAGGT TCGATGACCA TCAACTATCA GTTCGGGTGAT GTCGACGCTC 180
ACGGCGCCAT GATCCGCGCT CAGGCCGGGT TGCTGGAGGC CGAACATCAG GCCATCATTC 240
GTGATGTGTT GACCGCGAGT GACTTTTGGG GCGGCGCCGG TTCGGCGGCC TGCCAGGGGT 300
TCATTACCCA ATTGGGCCGT AACTTCCAGG TGATCTACGA ACAGGCCAAC GCCCACGGGC 360
AGAAGGTGCA GGCTGCCGGC AACAACATGG CGCAAACCGA CAGCGCCGTC GGCTCCAGCT 420
GGGCCTGACA CCAGGCCAAG GCCAGGGACG TGGTGTACGA GTGAAGGTTC CTCGCGTGAT 480
CCTTCGGGTG GCAGTCTAGG TGGTCAGTGC TGGGGTGTTG GTGGTTTGCT GCTTGGCGGG 540
TTCTTCGGTG CTGGTCAGTG CTGCTCGGGC TCGGGTGAGG ACCTCGAGGC CCAGGTAGCG 600
CCGTCCTTCG ATCCATTCGT CGTGTTGTTC GGCGAGGACG GCTCCGACGA GGCGGATGAT 660
CGAGGCGCGG TCGGGGAAGA TGCCCACGAC GTCGGTTCGG CGTCGTACCT CTCGGTTGAG 720
GCGTTCCTGG GGGTTGTTGG ACCAGATTTG GCGCCAGATC TTCTTGGGGA AGGCGGTGAA 780
CGCCAGCAGG TCGGTGCGGG CGGTGTCGAN GTGCTCGGCC ACCGCGGGGA GTTTGTCGGT 840
CAGAGCGTCG AGTACCCGAT CATATTGGGC AACAACTGAT TCGGCGTTGG GCTGGTCGTA 900

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1905 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
GCTCGCCGGA TGTGGGCGTC CGCGCAAAAC ATTTCCGGTG CGGGCTGGAG TGGCATGGCC      60
GAGGCGACCT CGCTAGACAC CATGGCCCAG ATGAATCAGG CGTTTCGCAA CATCGTGAAC     120
ATGCTGCACG GGGTGCGTGA CGGGCTGGTT CGCGACGGCA ACAACTACGA GCAGCAAGAG     180
CAGGCCTCCC AGCAGATCCT CAGCAGCTAA CGTCAGCCGC TGCAGCACAA TACTTTTACA     240
AGCGAAGGAG AACAGGTTCG ATGACCATCA ACTATCAGTT CGGTGATGTC GACGCTCACG     300
GCGCCATGAT CCGCGCTCAG GCCGGGTTGC TGGAGGCCGA GCATCAGGCC ATCATTCGTG     360
ATGTGTTGAC CGCGAGTGAC TTTTGGGGCG GCGCCGGTTC GGCGGCCTGC CAGGGGTTCA     420
TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAACA AGCCAACACC CACGGGCAGA     480
AGGTGCAAGC TGCCGGCAAC AACATGGCGC AAACCGACAG CGCCGTCNGC TCCAGCTGGC     540
CCTGACACCA GGCCAAGGCC AGGGACGTGG TGTACNAGTG AAGGTTCCTC GCGTGATCCT     600
TCGGGTGGCA GTCTAGGTGG TCAGTGCTGG GGTGTTGGTG GTTTGCTGCT TGGCGGGTTC     660
TTCGGTGCTG GTCAGTGCTG CTCGGGCTCG GGTGAGGACC TCGAGGCCCA GGTAGCGCCG     720
TCCTTCGATC CATTCGTCGT GTTGTTCGGC GAGGACNGCT CCGACGANGC GGATGATCGA     780
GGCGCGGTCG GGGAAGATGC CCACGACGTC GGTTCGGCGT CGTACCTCTC GGTTGAAGCG     840
TTCCTGGGGG CCACCGCTTG GCGCCNANGC ACTCCACGCC AATTCGTCNC ACCTAACAGC     900
GGTGGCCAAC GACTATGACT ACGACACCGT TTTTGCCAGG GCCCTCNAAA GGATCTGCGC     960
GTCCCGGCGA CACGCTTTTT GCGATAAGTA CCTCCGGCAA TTCTATGAGT GTACTGCGGN    1020
CCGCGAAAAC CGCAAGGGAG TTGGGTGTGA CGGTTNTTGC AAATGACGGG CGAATCCGGC    1080
GGCCAGCTGG CAGAATTCGC AGATTTCTTG ATCAACGTCC CGTCACGCGA CACCGGGCGA    1140
ATCCAGGAAT CTCACATCGT TTTTATTCAT GCGATCTCCG AACATGTCGA ACACGCGCTT    1200
TTCGCGCCTC GCCAATAGGA AAGCCGATCC TTACGCGGCC ATTCGAAAGA TGGTCGCGGA    1260
ACGTGCGGGA CACCAATGGT GTCTCTTCCT CGATAGAGAC GGGGTCATCA ATCGACAAGT    1320
GGTCGGCGAC TACGTACGGA ACTGGCGGCA GTTTGAATGG TTGCCCGGGG CGGCGCGGGC    1380
GTTGAAGAAG CTACGGGCAT GGGCTCCGTA CATCGTTGTC GTGACAAACC AGCAGGGCGT    1440
GGGTGCCGGA TTGATGAGCG CCGTCGACGT GATGGTGATA CATCGGCACC TCCAAATGCA    1500
GCTTGCATCC GATGGCGTGC TGATAGATGG ATTTCAGGTT TGCCCGCACC ACCGTTCGCA    1560
GCGGTGTGGC TGCCGTAAGC CGAGACCGGG TCTGGTCCTC GACTGGCTCG GACGACACCC    1620
CGACAGTGAG CCATTGCTGA GCATCGTGGT TGGGGACAGC CTCAGCGATC TTGACATTGG    1680
CACACAACGT CGCCGCTGCT GCCGGTGCAT GTGCCAGTGT CCAGATAGGG GGCGCCAGTT    1740
CTGGCGGTGT CGCTGACGCG TCATTTGACT CGCTCTGGGA GTTCGCTGTC GCAGTCGGAC    1800
ATGCGCGGGG GGAGCGGGGC TAATGGCGAT CTTGCGCGGG CGAGCGCCGT NGCGGNTCGG    1860
ACTNNGCGGT GGCGGGACAG ACGTGGAACC GTACTCGAGC CAGTT                    1905
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2921 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CGGGATGCCG TGGTGGTTGG TATTGCCCAA ACCCTGGCGC TGGTCCCCGG GGTATCCAGG      60
TCCGGGTCGA CCATCAGCGC TGGACTGTTT CTCGGACTCG ACCGTGAACT GGCCGCCCGA     120
TTCGGATTCC TGCTGGCCAT TCCAGCGGTG TTCGCCTCCG GGTTGTTCTC GTTGCCCGAC     180
GCATTCCACC CGGTAACCGA GGGCATGAGC GCTACTGGCC CGCAGTTGCT GGTGGCCACC     240
```

CTGATCGCGT TCGTCCTCGG TCTGACCGCG GTGGCCTGGC TGCTGCCGTT TCTGGTGCGA 300
CACAACATGT ACTGGTTCGT CGGCTACCGG GTGCTCGTCG GGACGGGCAT GCTCGTGCTG 360
CTGGCTACCG GGACGGTAGC CGCGACATGA CCGTCATCTT GCTACGCCAT GCCCGTTCCA 420
CCTCGAACAC CGCGGGCGTG CTGGCCGGCC GGTCCGGCGT CGACCTCGAC GAGAAGGGGC 480
GCGAGCAGGC CACCGGGTTG ATCGATCGAA TTGGTGACCT GCCGATCCGG GCGGTCGCGT 540
CTTCTCCAAT GCTGCGGTGT CAACGCACCG TCGAACCGCT GGCCGAGGCG CTGTGCCTGG 600
AGCCGCTCAT CGATGACCGG TTCTCCGAAG TCGACTACGG CGAATGGACT GGCAGAAAAA 660
TCGGTGACCT GGTCGACGAG CCGTTGTGGC GGGTAGTCCA GGCCCACCCC AGCGCGGCGG 720
TGTTTCCCGG CGGTGAGGGT TTGGCGCAGG TGCAGACGTG GTTGTCCTGA CGGATTTCCA 780
TGCCGGGGAA CACCAAGACC GGATCGGCAC TGGCGGTCGC CGGCGAAAAC CCGGCCGCCA 840
ATAGGGCGAC CGTCGCTGCG AATGCGCGTG GTACCAGGCG GACCACCTTG AACTCCCATC 900
CGTCGGGGCC AAGCGCATCG CCCGCCGCCG GTTACGGCTA AGGCGTACCA AAACCCGACG 960
GTAATACTTC GGCAATGTCG GGTCNCGACG TTACCGAGAC GTGACCAGNG AGGCNGCGGC 1020
ATTGGATTTA TCGATGGTGC GCGGTTCCCA NCCCGGCGGT CCGAANACGT AGCCCAGCCG 1080
ATCCCGCAGA CGTGTTGCCG ACCGCCAGTC ACGCACGATC GCCACGTACT CGCGGGTCTG 1140
CAGCTTCCAG ATGTTGAACG TGTCGACCCG CTTGGTCAGG CCATAATGCG GTCGGAATAG 1200
CTCCGGCTGA AAGCTACCGA ACAGGCGGTC CCAGATGATG AGGATGCCGC CATAGTTCTT 1260
GTCCANATAC ACCGGGTCCA TTCCGTGGTG GACCCGGTGG TGCGACGGGG TATTGAAGAC 1320
GAATTCGAAC CACCGCGGCA GCCTGTCGAT CCGCTCGGTG TGCACCCAGA ACTGGTAGAT 1380
CAAGTTCAGC GACCAATTGC AGAACACCAT CCAAGGGGGA AGCCCCATCA GTGGCAGCGG 1440
AACCCACATG AGAATCTCGC CGCTGTTGTT CCANTTTCTG GCGCAGCGCG GTGGCGAAGT 1500
TGAAGTATTC GCTGGAGTGA TGCGCCTGGT GGGTAGCCCA GATCAGCCGA ACTCGGTGGG 1560
CGATGCGGTG ATAGGAGTAG TACAGCAGAT CGACACCAAC GATCGCGATC ACCCAGGTGT 1620
ACCACCGGTG GGCGGACAGC TGCCAGGGGG CAAGGTAGGC ATAGATTGCG GCATAACCGA 1680
GCAGGGCAAG GGACTTCCAG CCGGCGGTGG TGGCTATCGA AACCAGCCCC ATCGAGATGC 1740
TGGCCACCGA GTCGCGGGTG AGGTAAGCGC CCGAGGCGGG CCGTGGCTGC CCGGTAGCAG 1800
CGGTCTCGAT GCTTTCCAGC TTGCGGGCCG CCGTCCATTC GAGAATCAGC AGCAATAGAA 1860
AACATGGAAT GGCGAACAGT ACCGGGTCCC GCATTTCCTC GGGCAGCGCT GAGAAGAATC 1920
CGGCGACGGC ATGGCCGAGG CGACCTCGNT AGACACCATG ACCCAGATGA ATCAGGCGTT 1980
TCGCAACATC GTGAACATGC TGCACGGGGT GCGTGACGGG CTGGTTCGCG ACGCCAACAA 2040
NTACGAACAG CAAGAGCAGG CCTCCCAGCA GATCCTCAGC AGCTGACCCG GCCCGACGAC 2100
TCAGGAGGAC ACATGACCAT CAACTATCAA TTCGGGGACG TCGACGCTCA CGGCGCCATG 2160
ATCCGCGCTC AGGCCGGGTC GCTGGAGGCC GAGCATCAGG CCATCATTTC TGATGTGTTG 2220
ACCGCGAGTG ACTTTTGGGG CGGCGCCGGT TCGGCGGCCT GCCAGGGGTT CATTACCCAG 2280
CTGGGCCGTA ACTTCCAGGT GATNTACGAG CAGGCCAACG CCCACGGGCA GAAGGTGCAG 2340
GCTGCCGGCA ACAACATGGC ACAAACCGAC AGCGCCGTCG GCTCCAGCTG GGCATAAAGN 2400
TGGCTTAAGG CCCGCGCCGT CAATTACAAC GTGGCCGCAC ACCGGTTGGT GTGTGGCCAC 2460
GTTGTTATCT GAACGACTAA CTACTTCGAC CTGCTAAAGT CGGCGCGTTG ATCCCCGGTC 2520
GGATGGTGCT GAACTGGGAA GATGGCCTCA ATGCCCTTGT TGCGGAAGGG ATTGAGGCCA 2580
TCGTGTTTCG TACTTTAGGC GATCAGTGCT GGTTGTGGGA GTCGCTGCTG CCCGACGAGG 2640
TGCGCCGACT GCCCGAGGAA CTGGCCCCGG TGGACGCATT GTTGGACGAT CCGGCGTTCT 2700
TCGCCCCGTT CGTGCCGTTC TTCGACCCGC GCAGGGGCCG GCCGTCGACG CCGATGGAGG 2760
TCTATCTGCA GTTGATGTTT GTGAAGTTCC GCTACCGGCT GGGCTATGAG TCGCTGTGCC 2820
GGGAGGTGGC TGATTCGATC ACCTGACGGC GGTTTTGCCG CATTGCGCTG GACGGGTCGG 2880
TGCCGCATCC GACCACATTG ATGAAGCTCA CCACGCGTTG C 2921

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1704 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8: